Serik Sagitov, Chalmers Tekniska Högskola, March 16, 2005

5. Limit theorems

5.1 Law of large numbers

Ex 1: coin tossing

Number of heads $X \sim \mathrm{Bin}(n, 1/2)$ in n tossings
proportion of heads $X/n \to 1/2$ as $n \to \infty$
since $\mathrm{E}(X/n) = 1/2$ and $\mathrm{Var}(X/n) = 1/(4n) \to 0$

Chebyshev's inequality

given $\mathrm{E}(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2$
$\mathrm{P}(|X - \mu| > \epsilon) \le \sigma^2/\epsilon^2$, $\forall \epsilon > 0$

Proof

$\mathrm{E}(1_{\{|X-\mu|>\epsilon\}}) \le \mathrm{E}(\frac{(X-\mu)^2}{\epsilon^2} 1_{\{|X-\mu|>\epsilon\}})$

Theorem LLN

let $\bar{X} = \frac{1}{n}(X_1 + \ldots + X_n)$ and $X_1, \ldots, X_n$ are
independent with $\mathrm{E}(X_i) = \mu$, $\mathrm{Var}(X_i) = \sigma^2$
then $\mathrm{P}(|\bar{X} - \mu| > \epsilon) \to 0$, $n \to \infty$, $\forall \epsilon > 0$

Proof

$\mathrm{Var}(\bar{X}) = \sigma^2/n$ apply Chebyshev's inequality

Ex 2: last ID digit

Collect the last ID digits in the class $X_1, \ldots, X_{30}$
ten sample counts $Y_0, Y_1, \ldots, Y_9$

Sample mean

$\bar{X} = \frac{1}{30}(X_1 + \ldots + X_{30}) = \frac{1}{30}(Y_1 + 2Y_2 + \ldots + 9Y_9)$

Ex 3: Monte-Carlo integration

Numerically compute $\int_0^1 e^{x^2} dx \approx \frac{1}{1000} \sum_{i=1}^{1000} e^{X_i^2}$
where $X_1, \dots, X_n$ are independent $\mathrm{U}(0,1)$
so that $\mathrm{E}(e^{X_i^2}) = \int_0^1 e^{x^2} dx$

5.2 Central Limit Theorem

Ex 2: last ID digit

$(X_i + 1) \sim \mathrm{U}(10)$, $\mathrm{Var}(X_i) = 99/12 = 8.25$
$\mathrm{Var}(\bar{X}) = 8.25/30 = 0.275 = (0.524)^2$
observe the difference $\bar{X} - 4.5$

Theorem CLT

let $S_n = X_1 + \ldots + X_n$ and $X_1, \dots, X_n$ are independent with $\mathrm{E}(X_i) = \mu$, $\mathrm{Var}(X_i) = \sigma^2$
then $\mathrm{P}(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le x) \to \Phi(x)$, $n \to \infty$, $\forall x$

Proof: mgf method, assume $\mu = 0$, $\sigma^2 = 1$
$M(t) = \mathrm{E}(e^{tX_i})$, $M(t) = 1 + \frac{1}{2}t^2 + o(t^2)$, $t \to 0$
$\mathrm{E}(e^{tS_n/\sqrt{n}}) = M^n(\frac{t}{\sqrt{n}}) \sim (1 + \frac{t^2}{2n})^n \to e^{t^2/2}$

Normal approximations

sample mean $\bar{X} \approx \mathrm{N}(\mu, \frac{\sigma^2}{n})$
$\mathrm{Bin}(n,p) \approx \mathrm{N}(np, npq)$, $np \ge 5$, $nq \ge 5$
$\mathrm{Pois}(\lambda) \approx \mathrm{N}(\lambda, \lambda)$, $\lambda \ge 5$
$\mathrm{Hg}(N,n,p) \approx \mathrm{N}(np, npq\frac{N-n}{N-1})$, $np \ge 5$, $nq \ge 5$
$\mathrm{Gamma}(\alpha, \lambda) \approx \mathrm{N}(\alpha/\lambda, \alpha/\lambda^2)$ for large α

Ex 4: diversification experiment

Three options of a special study support

a) take 4500 SEK

b) toss a coin and get 10000 SEK in case of heads

c) toss 10000 one-SEKs and collect all heads-up coins

Amount of money collected in the last case

$X \sim \text{Bin}(10000, 0.5)$, three-sigma rule: 5000 ± 150

Ex 5: aspirin teatment

$X = \#\{\text{heart attacks in the placebo group}\}$

Assuming no aspirin effect

$X \sim \text{Hg}(22071, 293, 0.4999) \approx \text{N}(146.48, 72.28)$

$\text{P}(X \geq 189) \approx 1 - \Phi(\frac{189-146.48}{8.50}) = 1 - \Phi(5)$

$= 0.0000003$ statistically significant aspirin effect

5.3 χ^2 and t distributions

Chi square distribution χ_1^2 with 1 degree of freedom

$Z^2 \sim \chi_1^2$, if $Z \sim \text{N}(0, 1)$

transformed r.v. $\Rightarrow \chi_1^2 = \text{Gamma}(1/2, 1/2)$

Chi square distribution with $k \geq 1$ degrees of freedom

$Z_1^2 + \ldots + Z_k^2 \sim \chi_k^2$, if independent $Z_i \sim \text{N}(0, 1)$

$\chi_k^2 = \text{Gamma}(k/2, 1/2)$, $\chi_2^2 = \text{Exp}(1/2)$

mgf $M(t) = (1 - 2t)^{-k/2}$

$f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{(k/2)-1} e^{-x/2}$, $\mu = k$, $\sigma^2 = 2k$

Sample mean and sample variance

Random sample from $\mathrm{N}(\mu, \sigma^2)$

independent r.v. $(X_1, \ldots, X_n)$, $X_i \sim \mathrm{N}(\mu, \sigma^2)$

sample mean $\bar{X} = \frac{X_1 + \ldots + X_n}{n} \sim \mathrm{N}(\mu, \frac{\sigma^2}{n})$

sample variance $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$

Theorem A

given a random sample from $\mathrm{N}(\mu, \sigma^2)$

$\bar{X}$ is independent of vector $(\bar{X} - X_1, \ldots, \bar{X} - X_n)$

Proof: mgf method, assume $\mu = 0$, $\sigma^2 = 1$

$\mathrm{E}\big(\exp\{s\bar{X} + \sum_{i=1}^n t_i(X_i - \bar{X})\}\big)$

$= \mathrm{E}\big(\exp\{\sum_{i=1}^n (sn^{-1} + t_i - \bar{t})X_i\}\big)$

$= \prod_{i=1}^n \exp\{(sn^{-1} + t_i - \bar{t})^2/2\}$

$= \exp\{\frac{s^2}{2n} + \frac{1}{2}\sum_{i=1}^n (t_i - \bar{t})^2\}$

$= \mathrm{E}(e^{s\bar{X}})\mathrm{E}\big(\exp\{\sum_{i=1}^n t_i(X_i - \bar{X})\}\big)$

Theorem B

random sample from $\mathrm{N}(\mu, \sigma^2) \Rightarrow (n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$

Proof: assume $\mu = 0$, $\sigma^2 = 1$

$$\underbrace{\sum_{i=1}^n X_i^2}_{\chi^2_n} = \underbrace{\sum_{i=1}^n (n\bar{X})^2}_{\chi^2_1} + \underbrace{\sum_{i=1}^n (X_i - \bar{X})^2}_{?}$$

apply Theorem A and mgf to show that

$\sum_{i=1}^n (X_i - \bar{X})^2 \sim \chi^2_{n-1}$

t-**distribution** with $k \geq 1$ degrees of freedom

$\frac{Z\sqrt{k}}{\sqrt{X}} \sim t_k$-distribution

if $Z \sim \mathrm{N}(0,1)$ and $X \sim \chi^2_k$ are independent

$f(t) = \frac{\Gamma((k+1)/2)}{\sqrt{k\pi}\,\Gamma(k/2)} \left(1 + \frac{t^2}{k}\right)^{-(k+1)/2}$

$$\boxed{\text{Random sample from } \mathrm{N}(\mu, \sigma^2) \Rightarrow \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}}$$

Normal approximation for the t-distribution

$t_k \approx \mathrm{N}(0,1)$ for large k