11. Sums of Independent Random Variables and Central Limit Theorem (P.483~P.535)

11.1 Moment-Generating Functions

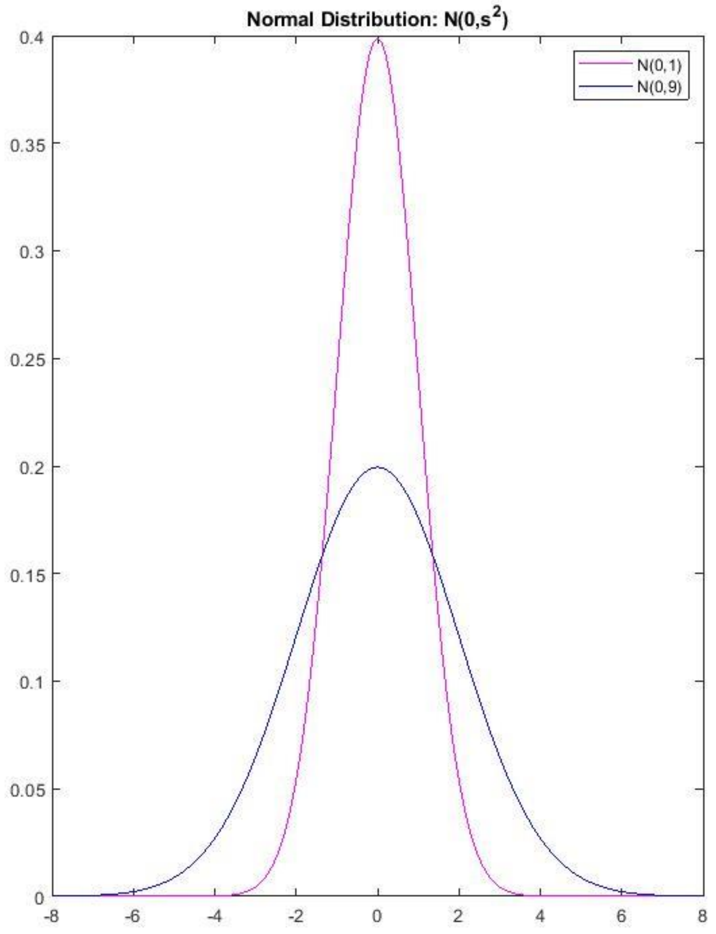
(11.1) For a random variable X, let $M_X(t) = E(e^{tX})$. If $M_X(t)$ is finite for all values of t in $(-\delta, \delta), \delta > 0,$ then $M_X(t)$ is called the moment-generating function of X.

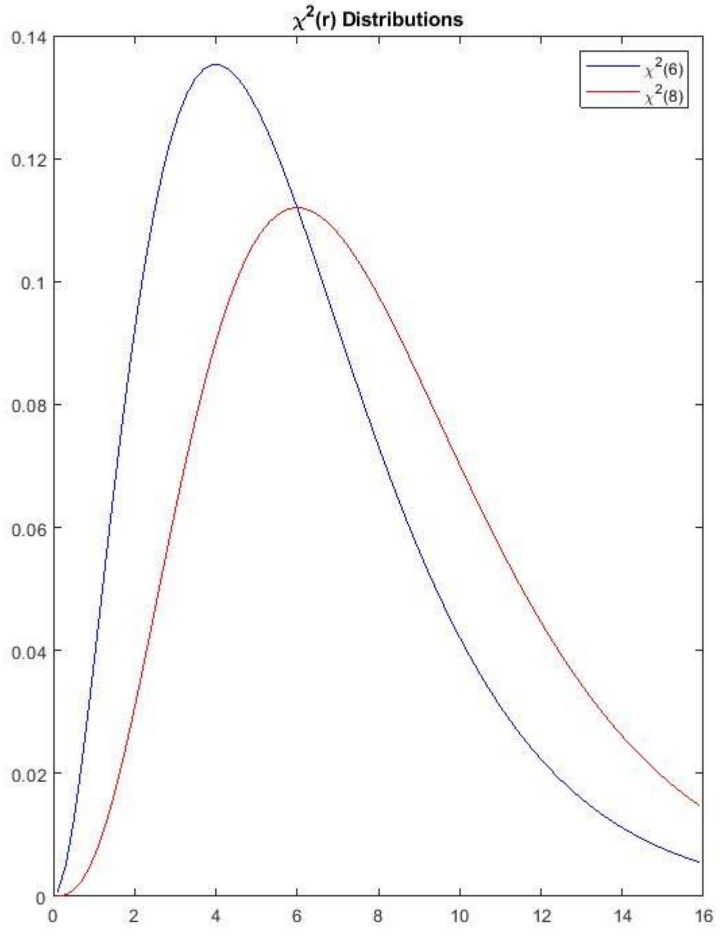
In a discrete case, $M_X(t) = E[e^{tX}] = \sum_{x \in A} e^{tx} p(x)$,

In a continuous case, $M_X(t) = \varphi_X(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f(x) dx$.

Let $\mu_X^{(r)} = E[(X - \mu)^r]$, the quantity $\mu_X^{(3)} / \sigma_X^3$, called the measure of skewness which is a measure of symmetry of the distribution function of X. Another quantity $\mu_X^{(4)} / \sigma_X^4$, called the measure of kurtosis which is a measure of relative flatness of the distribution function of X.

X~$N(0,\sigma^2)$ and X~$\chi^2(r)$





Theorem 11.1 $E(X^n) = M_X^{(n)}(0)$ on P.484

(Theorem 11.1) Let X be a r.v. with moment-generating function $M_X(t)$,

$M_X(t) = E[e^{tX}] = \sum_{x \in A} e^{tx} p(x)$, $M_X(t) = \varphi_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x)dx$.

Then $E(X^n) = M_X^{(n)}(0)$.

In particular, E(X)=$M_X'(0)$, Var(X)=$M_X''(0) - [M_X'(0)]^2$=$E(X^2)$-$[E(X)]^2$.

(Corollary) The MacLaurin's series for $M_X(t)$ is given by

$$M_X(t) = \sum_{n=0}^{\infty} \frac{M_X^{(n)}(0)}{n!} t^n = \sum_{n=0}^{\infty} M_X^{(n)}(0) \frac{t^n}{n!}.$$

Therefore, $E(X^n)$ is the coefficient of $\frac{t^n}{n!}$.

Moment-Generating Functions of Some Special Distributions

Compute the moment-generating functions of the following distributions.

(1) Bernoulli trial with parameter p, that is $P(X = 1) = p$.

(2) Binomial distribution $X{\sim}b(n, p)$.

(3) Poisson distribution with parameter λ.

(4) Geometric Distribution with parameter p.

(5) Exponential Distribution with parameter θ (or λ).

(6) Gamma Distribution with parameters $\alpha\ and\ \theta$.

(7) Normal Distribution $X{\sim}N(\mu, \sigma^2)$.

(8) $\chi^2(\mathrm{r})$, Chi-Square distribution with degrees of freedom r.

Theorem 11.2 on P.489

(Theorem 11.2) Let X and Y be two r.v.s with moment-generating functions $M_X(t)$ and $M_Y(t)$. If for some $\delta > 0, M_X(t)$=$M_Y(t)$ for all values of $t \in (-\delta, \delta)$, then X and Y have the same distribution.

(11.7) Let the moment-generating function of a r.v. X be

$M_X(t) = \frac{1}{7}e^t + \frac{3}{7}e^{3t} + \frac{2}{7}e^{5t} + \frac{1}{7}e^{7t}$. According to Theorem 11.2, the p.m.f. $p(x)$ is

$$p(1) = \frac{1}{7}, p(3) = \frac{3}{7}, p(5) = \frac{2}{7}, p(7) = \frac{1}{7}, and\ \ 0\ elsewhere.$$

(11.8) Let X be a r.v. with moment-generating function $M_X(t)$=e^{2t^2}.

Find P(0<X<1). (Ans:) $X \sim N(0, 2^2)$. $P(0 < X < 1) = P\left(0 < \frac{X}{2} < \frac{1}{2}\right) = \Phi(0.5) - \Phi(0) \approx 0.6915 - 0.5 = 0.1915.$

Theorems 11.3-11.5 on P.494-495

(11.3) Let $X_1, X_2, \cdots, X_n$ be independent r.v.s with moment-generating functions $M_{X_1}(t), M_{X_2}(t), \cdots, M_{X_n}(t)$. Then the moment-generating function of W$= \sum_{i=1}^{n} X_i$ is given by $M_W(t) = \prod_{i=1}^{n} M_{X_i}(t)$.

(11.4) Let $X_i \sim b(n_i, p), 1 \leq i \leq r$ *be independent*.

Define W$= \sum_{i=1}^{r} X_i$, then

$M_W(t)$=$(1 - p + pe^t)^{n_1+n_2+\cdots+n_r}$ and W$\sim b(n,p)$, *where* $n = \sum_{i=1}^{r} n_i$.

(11.5) Let $X_i \sim Poisson(\lambda_i), 1 \leq i \leq n$ *be independent*.

Define W$= \sum_{i=1}^{n} X_i$, then

$M_W(t) = \prod_{i=1}^{n} e^{\lambda_i(e^t-1)}$ and W$\sim Poisson(\lambda)$ *where* $\lambda = \sum_{i=1}^{n} \lambda_i$.

Theorems 11.6-11.7 on P.495-496

(11.6) Let $X_i \sim N\left(\mu_i, \sigma_i^2\right), 1 \leq i \leq n$ *be independent*.

Define W= $\sum_{i=1}^{n} X_i$, then $M_{X_i}(t) = e^{\mu_i t + \frac{\sigma_i^2 t^2}{2}}$, and $W \sim N(\mu, \sigma^2)$,
$M_W(t) = \prod_{i=1}^{n} M_{X_i}(t)$=$e^{\mu t + (\sigma^2 t^2/2)}$, where $\mu = \sum_{i=1}^{n} \mu_i$, $\sigma^2 = \sum_{i=1}^{n} \sigma_i^2$

(11.7) Let $X_i \sim N\left(\mu_i, \sigma_i^2\right), 1 \leq i \leq n$ *be independent*. Then

$\sum_{i=1}^{n} \alpha_i X_i \sim N(\sum_{i=1}^{n} \alpha_i \mu_i, \sum_{i=1}^{n} \alpha_i^2 \sigma_i^2)$. In particular, *the sample mean*
$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ with n independent $X_i \sim N(\mu, \sigma^2)$, *is* $N(\mu, \frac{\sigma^2}{n})$.

Example 11.9 on P.496-497

(11.9) Suppose that the distribution of students' grades in a probability test is normal with mean 72 and variance 25.

(a) What is the probability that the average grade of such a class with 25 students is 75 or more?

(b) If a Professor teaches two different sections of this course, each containing 25 students, what is the probability of one class is at least three more than the average of the other class?

Answer: Let $X_1, X_2, \cdots, X_{25}$ denote the grades of the 25 students. Then

$\{X_i \sim N(72, 5^2), i = 1, 2, \cdots, 25\}$ *is a random sample of size* 25.

(a) $\bar{X} \sim N\left(72, \frac{25}{25}\right), then\ P(\bar{X} \geq 75) = P(\bar{X} - 72 \geq 3) = 1 - \Phi(3) \approx 0.0013$

(b) $P(|\bar{X} - \bar{Y}| > 3) = 2P(\bar{X} - \bar{Y} > 3) = 2P\left(\frac{\bar{X} - \bar{Y} - 0}{\sqrt{2}} > \frac{3 - 0}{\sqrt{2}}\right) = 2(1 - \Phi(2.12)) \approx 0.034$

Examples on P.497-499

Recall a gamma probability density function, $(\lambda = \frac{1}{\theta})$.

$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, x > 0$, or $f(x) = \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-x/\theta}, x > 0.$

(A2) A gamma r.v. with parameters (n, λ) is the sum of n independent exponential random variables, each with mean $\frac{1}{\lambda}$, and vice versa.

(11.11) Let $X_i \sim N(0,1), 1 \le i \le n$ *be independent*. Define Y= X_1^2, then Y$\sim\chi^2(1)$, and $X = \sum_{i=1}^{n} X_i^2 \sim \chi^2(n)$, or say,

$$X \sim gamma\ with\ parameters\ (\frac{n}{2}, \frac{1}{2})$$

Problems on P.499

(A1) Let $X_1, X_2, \cdots, X_n$ be independent geometric random variables each with parameter p. Using moment-generating functions, prove that $W = \sum_{i=1}^{n} X_i$ is negative binomial with parameters (n, p).

(Ans:) $M_i(\mathrm{t}) = \frac{pe^t}{1-(1-p)e^t}$ for $random\ variable\ X_i$, $i = 1,2,\cdots,n$.

$$M_W(t) = \frac{(pe^t)^n}{(1-(1-p)e^t)^n},$$

Thus, W is a random variable of negative binomial with parameters (n, p).

Problems on P.499

(A9) Let $X \sim N(1,2)$ and $Y \sim N(4,7)$ be independent r.v.s. Then

$$(a)\ X+Y \sim N(5,9), and\ P(X+Y>0) = P\left(\frac{X+Y-5}{3} > \frac{0-5}{3}\right)$$

$$= 1-\Phi(-1.67) = 0.9525$$

$$(b)\ X-Y \sim N(-3,9), and\ P(X-Y<2) = P\left(\frac{X-Y+3}{3} < \frac{2+3}{3}\right)$$

$$= \Phi(1.67) = 0.9525.$$

$$(c)\ 3X+4Y \sim N(19,130), and\ P(3X+4Y>20) = P(\tfrac{3X+4Y-19}{\sqrt{130}} > \tfrac{20-19}{\sqrt{130}})$$

$$= 1-\Phi(0.0877) \approx 0.4671$$

Markov's Inequality on P.502

(11.8) (Markov's Inequality)

Let X be a nonnegative random variable, then for any $t > 0$,

$$P(X \geq t) \leq \frac{E(X)}{t}.$$

Proof: Let A be the set of possible values of X and $B = \{x \in A: x \geq t\}$.

Then

$$E(X) = \sum_{x \in A} xp(x) \geq \sum_{x \in B} tp(x) = tP(X \geq t), thus\ P(X \geq t) \leq \frac{E(X)}{t}.$$

(11.12) A post office, on average, handles 10,000 letters per day. What can be said about probability that it will handle (a) at least 15,000 letters tomorrow; (b) fewer than 15,000 letters tomorrow?

Ans: Let X be the number of letters that this office will handle tomorrow. Then

E(X)=10000, and (a) P(X$\geq$ 15000) $\leq \frac{E(X)}{15000}$ =2/3, (b) P(X<15000)=1-P(X$\geq$ 15000) $\geq$1/3.

Chebyshev's Inequality on P.502

(11.9) (Chebyshev's Inequality)

If X is a random variable with *expected value* μ *and variance* σ^2, then for any $t > 0$,

$$P(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2}.$$

Proof: By Markov's inequality, $P(|X - \mu|^2 \geq t^2) \leq \frac{E[(X-\mu)^2]}{t^2} = \frac{\sigma^2}{t^2}$

Letting $t = k\sigma$, then we have $P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$

Examples 11.13-11.14 on P.503-504

(11.13) Suppose that, on average, a post office handles 10,000 letters a day with a variance of 2000. What can be said about the probability that this post office will handle between 8000 and 12,000 letters tomorrow?

Ans: $P(8000 \leq X \leq 12000) = P(|X - 10000| \leq 2000)$

$= 1 - P(|X - 10000| > 2000) \geq 1 - \frac{2000}{(2000)^2}$ =0.9995

(11.14) A blind will fit Myra's bedroom's window if its width is between 41.5 and 42.5 inches. Myra buys a blind from a store that has 30 such blinds. What can be said about the probability that it fits her window if the average of the widths of the blinds is 42 inches with standard deviation 0.25?

Ans: Let X be the the width of the blind that Myra purchased, then

we have $P(41.5 < X < 42.5) = P(|X - 42| < 2(0.25)) \geq 1 - \frac{1}{2^2} = 0.75$, where σ^2=0.25.

Chebyshev’s Inequality and Sample Mean (P.505)

- Let $X_1, X_2, \cdots, X_n$ be a random sample of size n with distribution function F and mean μ, variance σ^2. Then
- $E(\bar{X})$=E($\frac{X_1+X_2+\cdots+X_n}{n}$) $= \mu$,
- $Var(\bar{X}) = Var(\frac{X_1+X_2+\cdots+X_n}{n})$=$\frac{\sigma^2}{n}$.
- $P(|\bar{X} - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$

(11.17) For the scores on an achievement given to a certain population of students, the expected value is 500 and the standard deviation is 100. Let $\bar{X}$ be the mean of the scores of a random sample of 10 students. Find a lower bound for $P(460 < \bar{X} < 540)$.

Ans: $P(460 < \bar{X} < 540) = P(|X - 500| < 40) = 1 - P(|X - 500| \geq 40)$

$$\geq 1 - \frac{(100)^2}{(40)^2 \times 10} = 1 - \frac{5}{8} = \frac{3}{8}$$

11.4 Laws of Large Numbers on P.511-513

(11.10) Weak Law of Large Numbers

Let $X_1, X_2, \cdots, X_n$ be a sequence of independent and identically distributed random variables with $\mu = E(X_i), and\ \sigma^2 = Var(X_i) < \infty$,

Then $\forall \varepsilon > 0$,

$$\lim_{n\to\infty} P(|\frac{\sum_{i=1}^{n} X_i}{n} - \mu| > \epsilon) = 0.$$

(11.12) Strong Law of Large Numbers

Let $X_1, X_2, \cdots, X_n$ be a sequence of independent and identically distributed random variables with $\mu = E(X_i), i = 1,2, \cdots n$. Then

$$P([\lim_{n\to\infty} \frac{X_1 + X_2 + \cdots + X_n}{n}] = \mu) = 1$$

~~Examples 11.21-11.26 on P.514-518~~

~~(11.21) on P.514~~

~~(11.22) on P.514-P.515~~

~~(11.23) on P.515-P.516~~

~~(11.24) on P.516~~

~~(11.25) on P.516-P.517~~

~~(11.26) on P.517-P.518~~

11.5 Central Limit Theorem on P.520

(Theorem 11.12) Central Limit Theorem

Let $X_1, X_2, \cdots, X_n$ be a sequence of independent and identically distributed random variables with $\mu = E(X_i), and\ \sigma^2 = Var(X_i) < \infty$,

Then the distribution of

$$Z_n = \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

converges to the distribution of a standard normal random variable. That is,

$$\lim_{n\to\infty} P(Z_n \le x) = \lim_{n\to\infty} P\left(\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \le x\right)$$

$$= \lim_{n\to\infty} P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \le x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2}\, dy$$

Proof of Central Limit Theorem

- Provided in class Website
- http://www.cs.nthu.edu.tw/~cchen/EECS3030/eecs3030.html

From previous slide (P.8) in this Section,

Let $\{X_i \sim N(\mu, \sigma^2), 1 \leq i \leq n\}$ *be a random sample*, $M_{X_i}(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$,

Define W= $\sum_{i=1}^{n} X_i$, then $M_W(t) = e^{n\mu t + \frac{n\sigma^2 t^2}{2}}$, $W \sim N(n\mu, (\sqrt{n}\sigma)^2)$,

Thus,

$$\frac{W - n\mu}{\sqrt{n}\sigma} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

Examples 11.27-11.28 on P.522-P.523

(11.27) The lifetime of a TV tube (in years) is an exponential random variable with mean 10. What is the probability that the average lifetime of a random sample of 36 TV tubes is at least 10.5?

Ans: $P(\bar{X} \geq 10.5) = P\left(\frac{\bar{X}-10}{10/\sqrt{36}} \geq \frac{10.5-10}{10/\sqrt{36}}\right) = 1 - \Phi(0.30) \approx 1 - 0.6179 \approx 0.3821.$

(11.28) The lifetime it takes for a student to finish an aptitude test (in hours) has the probability density function

$$f(x) = 6(x-1)(2-x) \; if \; 1 < x < 2.$$

Approximate the probability that the average length of time it takes for a random sample of 15 students to complete the test is less that 1 hour 25 minutes (85 min).

Ans: $E(X_i) = \frac{3}{2} = 1.5, \; E\left(X_i^2\right) = \frac{23}{10} = 2.3, \; Var(X_i) = \sigma^2 = \frac{1}{20}, \; \sigma \approx 0.2236,$

$P\left(\bar{X} < \frac{85}{60}\right) = \mathrm{P}\left(\frac{\bar{X}-1.5}{\frac{0.2236}{\sqrt{15}}} < \frac{\frac{17}{12}-1.5}{\frac{0.2236}{\sqrt{15}}}\right) = P\left(\frac{\bar{X}-1.5}{\frac{0.2236}{\sqrt{15}}} < -1.44\right) \approx \Phi(-1.44) \approx 0.0749.$

Example 11.29 on P.523

(11.29) If 20 random numbers are selected independently from the interval (0, 1), what is the approximate probability that the sum of those numbers is at least 8?

Ans: $P(\sum_{i=1}^{20} X_i \geq 8) = P\left(\frac{\sum_{i=1}^{20} X_i - 20(\frac{1}{2})}{\sqrt{1/12}\sqrt{20}} \geq \frac{8-20(\frac{1}{2})}{\sqrt{1/12}\sqrt{20}}\right) \approx P(Z \geq -1.55)$

$$= 1 - \Phi(-1.55) \approx 0.9394.$$

Example 11.30 on P.523-P.524

(11.30) A biologist wants to estimate l, the life expectancy of a certain type of insect. To do so, he takes a sample of size n and measures the lifetime from birth to death of each insect. Then he finds the average of these numbers. If he believes that the lifetimes of these insects are independent random variables with variance *1.5 days*, how large a sample should he choose to be 98% sure that this average is accurate within $\pm 4.8\ hours\ (\pm 0.2\ days)$?

Ans: $P\left(-0.2 < \frac{X_1+X_2+\cdots+X_n}{n} - l < 0.2\right) \approx 0.98.$

Since $E(X_i) = l,\ Var(X_i) = 1.5$, by the central limit theorem,

$$P\left(-0.2n < \sum_{i=1}^{n} X_i - nl < 0.2n\right) = \mathrm{P}\left(-\frac{0.2n}{\sqrt{1.5n}} < \frac{\sum_{i=1}^{n} X_i - nl}{\sqrt{1.5n}} < \frac{0.2n}{\sqrt{1.5n}}\right)$$

$= \Phi\left(\frac{0.2n}{\sqrt{1.5n}}\right) - \Phi\left(-\frac{0.2n}{\sqrt{1.5n}}\right) = 2\Phi\left(\frac{0.2\sqrt{n}}{\sqrt{1.5}}\right) - 1 \approx 0.98$, thus $n \approx 203.58 \approx 204$.

A Simulation for $Z \sim N(0,1)$ By C.L.T.

- Let $\{X_1, X_2, \cdots, X_n\}$ be a random sample from U(0,1). Then
- $E(X_i) = \frac{1}{2}, and\ Var(X_i) = \frac{1}{12}$. Define $Z = \frac{X_1+X_2+\cdots+X_n-0.5n}{\sqrt{n/12}}$
- By the central limit theorem,
  $Z\ approximates\ the\ standard\ normal\ distribution$.
- To simulate the standard normal distribution Z, we can choose *n=12*,
- get a z value by computing $z = x_1 + x_2 + \cdots + x_{12} - 6$,
- $where\ each\ \ x_i$ could be obtained from a system built-in random().