Probability I (B) 97

Chapter 7 : Normal distribution and Central limit theorem

National Taiwan Normal University. *Lecturer: Pi-Wen Tsai*

Normal distribution

A r.v. is a normal (written $\mathcal{N}(\mu, \sigma^2)$) if it has density

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

A synonym for normal is Gaussian.

A r.v. is a standard normal (written $\mathcal{N}(0, 1)$) if it has density

$$\frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Theorem 7-1 (Thm 5.2-1)**.** *If X is $\mathcal{N}(\mu, \sigma^2)$ and $Z = (X - \mu)/\sigma$, then Z is $\mathcal{N}(0, 1)$.*

The distribution function of a standard $\mathcal{N}(0, 1)$ is often denoted $\Phi(x)$, so that

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} dy.$$

Tables of $\Phi(x)$ are often given only for $x > 0$. One can use the symmetry of the density function to see that $\Phi(-x) = 1 - \Phi(x)$.

Moment generating function for a normal R.V.

Distributions from Normal distribution

Theorem 7-2 (Thm 5.2-2)**.** $Z^2 \sim \chi^2(1)$ *distribution*
If Z is a standard normal random variable, the distribution of $U = Z^2$ is called the chi-square distribution with 1 degree of freedom.

$$\begin{aligned}F_Y(x) &= \mathsf{P}(Y \le x) = \mathsf{P}(X^2 \le x) = \mathsf{P}(-\sqrt{x} \le X \le \sqrt{x})\\ &= \mathsf{P}(X \le \sqrt{x}) - \mathsf{P}(X \le -\sqrt{x}) = F_X(\sqrt{x}) - F_X(-\sqrt{x}).\end{aligned}$$

Taking the derivative and using the chain rule,

$$f_Y(x) = \frac{d}{dx}F_Y(x) = f_X(\sqrt{x})\Big(\frac{1}{2\sqrt{x}}\Big) - f_X(-\sqrt{x})\Big(-\frac{1}{2\sqrt{x}}\Big).$$

Remembering that $f_X(t) = \frac{1}{\sqrt{2\pi}}e^{-t^2/2}$ and doing some algebra, we end up with

$$f_Y(x) = \frac{1}{\sqrt{2\pi}}x^{-1/2}e^{-x/2},$$

which is a Gamma with parameters $\frac{1}{2}$ and $\frac{1}{2}$. (This is also a χ^2 with one degree of freedom.)

Note that $\chi^2(1)$ is a special case of gamma distribution with parameters $\frac{1}{2}$ and $\frac{1}{2}$,

Theorem 7-3 (Thm 5.3-2)**.** $\chi^2(n)$ *distribution*
If U_i are independent chi-square RVs with 1 df, the distribution of $V = \sum_{i=1}^n U_i$ is a chi-square distribution with n degrees of freedom.

Definition 7-4 (ex 5.3-5)**.** Student t distribution
If $Z \sim \mathcal{N}(0,1)$, $U \sim \chi^2(v)$ and Z and U are independent, then the distribution of

$$\frac{Z}{\sqrt{U/v}}$$

is called the **t distribution** with v degree of freedom and is denoted by $t(v)$

The density function of the t distribution with v degree of freedom is

t is symmetric about zero.

heavier tail than a standard normal

$t \to \mathcal{N}(0,1)$ as $n \to \infty$.

Definition 7-5 (ex 4,4-5 p 235). $F(m, n)$ distribution

If U and V are independent chi-square RVs with m and n d.f. respectively, then the distribution of

$$\frac{U/m}{V/n}$$

is called the **F distribution** with m and n degrees of freedom, and is denoted by $F(m, n)$.

If $X \sim F(n, m)$, then $X^{-1} \sim F(m, n)$

If $T \sim t(n)$, then $T^2 = F(1, n)$

Random Sample for a population

The X_i constitute a random sample of size n when X_i are independent and the X_i are identically distributed, that is each X_i comes from the same distribution $f_X(x_i)$. We say X_i are "i.i.d".

Note that, uppercases letters, $X_1, \cdots, X_n$, denote the definition of the random variable and the lowercase letters, $x_1, \cdots, x_n$, denote the realized value of the observed random variable.

The mean and variance of $\overline{X}$ of n independent observations are $\mu_{\overline{X}} = \mu$ and $\sigma^2_{\overline{X}} = \frac{\sigma^2}{n}$

The sampling distribution of $\overline{X}$ from a normal population

Theorem 7-6 (Thm5.3-1). *If the underlying population itself has the $\mathcal{N}(\mu, \sigma^2)$ distribution, then the sample means $\overline{X}$ of n independent observations has the $\mathcal{N}(\mu, \frac{\sigma^2}{n})$ distribution.*

Moment generating function for $\overline{X}$

Note: any linear combination of **independent normal** RVs is also normally distributed.

Theorem 7-7 (Thm 5.3-3). *X_i are r.s. from a normal population, let $\overline{X} = \frac{X_1 + \cdot + X_n}{n}$ and $S^2 = \frac{\sum(X_i - \overline{X})^2}{n-1}$, then*

- *$\overline{X}$ and S^2 are independent*
- *$\frac{(n-1)S^2}{\sigma^2}$ is a $\chi^2(n-1)$ distribution*

Proof.

If X_i is a $\mathcal{N}(\mu_i, \sigma_i^2)$ and the X_i are independent, then $\sum_{i=1}^n c_i X_i$ is a $\mathcal{N}(\sum c_i\mu_i, \sum c_i^2\sigma_i^2)$.

The distribution of $X+Y$ when X and Y are independent Normal random variables with parameter (μ_1, σ_1^2) and (μ_2, σ_2^2), respectively?

$\mathcal{N}(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$

If X and Y are independent normals, then $-Y$ is also a normal with $\mathbb{E}(-Y) = -\mathbb{E}Y$ and $\mathrm{Var}(-Y) = (-1)^2\mathrm{Var}Y = \mathrm{Var}Y$, and so $X-Y$ is also normal.

$\mathcal{N}(\mu_1-\mu_2, \sigma_1^2+\sigma_2^2)$

Central Limit Theorem CLT

Definition 7-8 ((Weak) Law of Large number, WLLN)**.** Let $X_1, X_2, \cdots$ be a sequence of independent RVs with a mean μ and a variance σ^2 which are finite. Let $\overline{X}_n = n^{-1}\sum X_i$ be the sample mean. Then, for every $\epsilon > 0$,

$$\lim_{n\to\infty} \mathsf{P}(|\overline{X}_n - \mu| \le \epsilon) = 1$$

or, equivalently,

$$\lim_{n\to\infty} \mathsf{P}(|\overline{X}_n - \mu| > \epsilon) = 0,$$

sometime write as $\mathsf{P}(|\overline{X}_n - \mu| < \epsilon) \to 1$ as $n \to \infty$

Definition 7-9 (Convergence in probability)**.** If a sequence of random variable $\{Z_n\}$, is such that for every $\epsilon > 0$,

$$\lim_{n\to\infty} \mathsf{P}(|Z_n - a| \le \epsilon) = 1$$

where a is some scalar, then Z_n is said to converge in probability to a.

Z_n sometimes converges in probability to another RV, $Z_n \xrightarrow{p} Z$.

WLLN can be say as $\overline{X}_n$ converges in probability to μ, and sometimes write as $\overline{X}_n \xrightarrow{p} \mu$.

Definition 7-10 (Convergence in distribution)**.** Let $X_1, X_2, \cdots$ be a sequence of RVs with cumulative distribution functions $F_1, F_2, \cdots,$. Let X be a RV with cumulative distribution function F_X. We say the X_n converges in distribution to X if

$$\lim_{n\to\infty} F_n(x) = F_X(x)$$

at every point at which F_X is continuous. Sometimes write as $X_n \xrightarrow{d} X$.

Theorem 7-11 (Thm5.4-1, Central limit theorem (CLT))**.** *Let $X_1, X_2, \cdots, X_n$ be a random sample of size n from a population with a mean μ and a variance σ^2 which are finite. Let $\overline{X}$ be the sample mean. Then the distribution of*

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

tends to the standard normal as $n \to \infty$. That is,

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1).$$

(Limiting distribution)

Asymptotic distributions

For large sample size n, the distribution of $W_n = \sqrt{n}(\overline{X} - \mu)/\sigma$ gets closer to a standard normal no matter what shape the underlying population distribution has, as long as the population has a finite variance.

$$W_n \overset{A}{\sim} \mathcal{N}(0, 1).$$

Or the distribution of $\overline{X}$ gets closer to a normal distribution with mean μ variance σ^2/n as n is large.

$$\overline{X} \overset{A}{\sim} \mathcal{N}(\mu, \frac{\sigma^2}{n})$$

The normal approximations for discrete distributions

The normal approximation to the Binomial Distribution

For n independent trials with success probability p

$$\mathsf{P}(a \text{ to } b \text{ successes }) \approx \Phi(\frac{b + 1/2 - \mu}{\sigma}) - \Phi(\frac{a - 1/2 - \mu}{\sigma})$$

where $\mu = np$ is the mean and $\sigma = \sqrt{npq}$ is the standard deviation.

P(20 heads in 40 fair flips)

pnorm(20.5, 20, sqrt(10))-pnorm(19.5,20,sqrt(10))

0.1256329

dbinom(20, 40, .5)

0.1253707

P(more than 150 students within the 450 applications when it is known on average only 30 percent applications will really come)

The normal approximation to the Poisson Distribution

Bivariate Normal distribution

Bivariate Normal distribution

bivariate normal pdf

marginal distributions of a bivariate normal are again normal

conditional distributions of a bivariate normal are again normal

Theorem 7-12 (Thm 5.6-1). *X and Y have a bivariate normal distribution, then X and Y are independent if and only if $\rho = 0$.*