Recap:

① Convergence concepts for a sequence of random variables.

(i) Convergence in prob.

$X_1, \ldots, \ldots$ converges in prob. to X, denoted by $X_n \xrightarrow{P} X$ if for every $\varepsilon > 0$,

$$\lim_{n \to \infty} P(|X_n - X| > \varepsilon) = 0$$

(ii) Convergence in distribution:

$X_1, \ldots$ converges in dist. to X, denoted by $X_n \xrightarrow{d} X$ if

$$\lim_{n \to \infty} F_{X_n}(x) = F_X(x), \text{ for all continuity points } x \text{ of } F_X.$$

F_{X_n} = CDF of X_n

F_X = CDF of X.

(iii) Convergence in prob. $\Rightarrow$ Convergence in dist.
$\not\Leftarrow$ in general.

Slutskys theorem:

If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{P} a$ (a is just a number),

then a) $Y_n X_n \xrightarrow{d} aX$ b) $Y_n + X_n \xrightarrow{d} X + a$

Weak law of large numbers

If $X_1 \cdots\cdots$ are i.i.d random variables

with $E[X_i] = \mu$, $Var(X_i) = \sigma^2 < \infty$ then,

$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow{P} \mu$

Example: Toss a coin repeatedly.

$X_i = 1$ if the ith toss results in head

$= 0$ o.w.

$X_i \sim Ber(\mu)$ $E[X_i] = \mu$ $Var(X_i) = \mu(1-\mu) < \infty$

By weak law of large number $\frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow{P} \mu$

$\frac{1}{n} \sum_{i=1}^{n} X_i$ = average number of heads in n ~~trials~~ tosses.

Central limit theorem:

Let $X_1 \cdots\cdots$ be a sequence of random variables which are i.i.d and whose MGF is bounded near 0. Let $E[X_i] = \mu$, $Var(X_i) = \sigma^2 > 0$,

Define $\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$, Let $Z_n = \sqrt{n} \frac{(\overline{X}_n - \mu)}{\sigma}$

then $Z_n \xrightarrow{d} N(0,1)$

If we have a sequence of random variables $X_1 \ldots\ldots$ s.t. $\sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \xrightarrow{d} N(0,1)$

or equivalently, $\sqrt{n}(\bar{X}-\mu) \xrightarrow{d} N(0, \sigma^2)$

then, is it possible to find the asymptotic distribution of $\sqrt{n}\left(g(\bar{X}_n) - g(\mu)\right)$

where g is any function?

Delta theorem:

Let Y_n be a sequence of random variables that satisfies $\sqrt{n}(Y_n - \theta) \xrightarrow{d} N(0, \sigma^2)$. For a given function g, ~~and~~ suppose $g'(\theta)$ exists and is not 0. Then

$$\sqrt{n}\left(g(Y_n) - g(\theta)\right) \xrightarrow{d} N(0, \sigma^2 [g'(\theta)]^2)$$

If $g'(\theta) = 0$ and $g''(\theta)$ exists and is nonzero,

then $n\left(g(Y_n) - g(\theta)\right) \xrightarrow{d} \sigma^2 \frac{g''(\theta)}{2} \chi^2_1$

Intuition: $g(x) \approx g(\theta) + (x-\theta)\, g'(\theta)$

$g(Y_n) \approx g(\theta) + (Y_n - \theta)\, g'(\theta)$

$\Rightarrow \sqrt{n}\left(g(Y_n) - g(\theta)\right) \simeq \underbrace{\sqrt{n}(Y_n - \theta)}\, g'(\theta)$

$\sqrt{n}(Y_n - \theta) \sim N(0, \sigma^2)$ for large n approximately

$\Rightarrow \sqrt{n}\left(g(Y_n) - g(\theta)\right) \sim N(0, [g'(\theta)]^2 \sigma^2)$ for large n approximately.

Example: $X_1 \cdots$ are iid Ber(μ)

$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \mu(1-\mu))$ by CLT.

Qn: Find the approximate dist. of $\sqrt{n}(\bar{X}_n^2 - \mu^2)$

$g(x) = x^2 \quad g'(x) = 2x, \quad g'(\mu) = 2\mu \neq 0$

$\sqrt{n}(g(\bar{X}_n) - g(\mu)) \xrightarrow{d} N(0, \mu(1-\mu)[g'(\mu)]^2)$

$\Rightarrow \sqrt{n}(\bar{X}_n^2 - \mu^2) \xrightarrow{d} N(0, \mu(1-\mu)(2\mu)^2)$

Example: $X_1 \cdots$ are iid Ber$(\frac{1}{2})$

Qn: Find distn. of $\bar{X}_n(1-\bar{X}_n)$

~~g(x) = x~~ I know, $\sqrt{n}(\bar{X}_n - \frac{1}{2}) \xrightarrow{d} N(0, \frac{1}{4})$

$g(x) = x(1-x)$

$\Rightarrow g'(x) = 1-2x \quad \Rightarrow g'(\frac{1}{2}) = 0$ at $x = \frac{1}{2}$

however, $g''(x) = -2$

hence, $n\left(\bar{X}_n(1-\bar{X}_n) - \frac{1}{2}\cdot(1-\frac{1}{2})\right) \xrightarrow{d} \frac{1}{4}\cdot\frac{(-2)}{2}^2 \chi_1^2$

Statistical Inference:

$X_1, \cdots, X_n$ is a random sample.

$X_1, \cdots, X_n \overset{iid}{\sim} f(x|\theta)$.

Goal: To estimate θ.

provide a point estimate of θ.

We reduce the data in our real life applications, i.e. we start with $(X_1, \dots, X_n)$ a n-variate data vector and we end up using $T(X_1, \dots, X_n)$ which is a smaller variate function.

For example

$$T(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i, \quad \text{or} \quad T(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$$

~~Qn. If $X_1 \dots X_n$~~

Qn: $\underline{X} = (X_1, \dots, X_n) \sim F(\underline{x} \mid \theta)$. $T(\underline{X})$ is a function of $\underline{X}$. How much reduction of $\underline{X}$ is allowable without losing any "information" on θ?

Sufficiency principle:

Definition 1: Let $\underline{X} = (X_1, \dots, X_n)$ and $\underline{X} \sim F(\underline{x} \mid \theta)$. $T(\underline{X})$ is known to be a sufficient statistic for θ if the conditional distribution of $\underline{X} \mid T(\underline{X})$ is independent of θ.

Intuitively, $T(\underline{X})$ contains the same information about θ that $\underline{X}$ contains.

Example: $X_1, X_2, X_3 \overset{iid}{\sim} \text{Ber}(p)$.

$$P(X_1 = x) = p^x (1-p)^{1-x}, \text{ if } x = 0, 1.$$

$$T(X_1, X_2, X_3) = X_1 + X_2 + X_3$$

Lets look at the distribution of $(X_1, X_2, X_3) \,|\, X_1 + X_2 + X_3$.

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3 \,|\, X_1 + X_2 + X_3 = t)$$

$= 0$ if $t \neq x_1 + x_2 + x_3$

when $t = x_1 + x_2 + x_3$

$$= \frac{P(X_1 = x_1, X_2 = x_2, X_3 = x_3, X_1 + X_2 + X_3 = t)}{P(X_1 + X_2 + X_3 = t)}$$

$$= \frac{P(X_1 = x_1, X_2 = x_2, X_3 = x_3)}{P(X_1 + X_2 + X_3 = t)}$$

$$= \frac{P(X_1 = x_1)\, P(X_2 = x_2)\, P(X_3 = x_3)}{P(X_1 + X_2 + X_3 = t)}$$

$$= \frac{p^{x_1}(1-p)^{1-x_1}\, p^{x_2}(1-p)^{1-x_2}\, p^{x_3}(1-p)^{1-x_3}}{P(X_1 + X_2 + X_3 = t)}$$

$$= \frac{p^{x_1+x_2+x_3}(1-p)^{3-x_1-x_2-x_3}}{P(X_1 + X_2 + X_3 = t)}$$

$$= \frac{p^{x_1+x_2+x_3}(1-p)^{3-(x_1+x_2+x_3)}}{\binom{3}{t} p^t (1-p)^{3-t}}$$

$$= \frac{p^t \cancel{(1-p)^{3-t}}}{\binom{3}{t} p^t \cancel{(1-p)^{3-t}}} = \frac{1}{\binom{3}{t}}$$

Thus $(X_1, X_2, X_3) \mid X_1+X_2+X_3$ is independent of p.

Hence by definition of sufficient statistic $X_1+X_2+X_3$ is a sufficient statistic for p.

$X_1, X_2, X_3 \overset{iid}{\sim} Ber(p)$

	cases	prob.
$X_1+X_2+X_3=0 \rightarrow$	000	$(1-p)^3$
	001	$(1-p)^2 p$
	100	$p(1-p)^2$
$X_1+X_2+X_3=1 \rightarrow$	010	$p(1-p)^2$
	110	$p^2(1-p)$
$X_1+X_2+X_3=2 \rightarrow$	011	$p^2(1-p)$
	101	$p^2(1-p)$
$X_1+X_2+X_3=3 \rightarrow$	111	p^3

Therefore, $\Omega = \{000, 001, 100, 010, 110, 011, 101, 111\}$

$= A_0 \cup A_1 \cup A_2 \cup A_3$

$A_0 = \{000\}$, $A_1 = \{001, 100, 010\}$,

$A_2 = \{110, 011, 101\}$, $A_3 = \{111\}$

within A_t, $X_1+X_2+X_3 = t$.

and within each A_t for any sample point the prob. is $p^t(1-p)^{3-t}$

Goal: Find sufficient statistic from any situation where joint density / p.m.f. of $\underline{X} = (X_1, \ldots, X_n)$ is known.

Theorem (Factorization theorem):

Let $\underline{X}$ have joint p.d.f. (or p.m.f.) $f_\theta(\underline{x})$, where θ is the unknown parameter. A statistic $T(\underline{X})$ is sufficient for θ if and only if $f_\theta(\underline{X})$ can be expressed as

$f_\theta(\underline{X}) = g(T(\underline{X}), \theta)\, h(\underline{X})$, when $h(\underline{X})$ is a function of $\underline{X}$ independent of θ.

proof: only if part:

$$P(\underline{X} = \underline{x}) = \sum_t P(\underline{X} = \underline{x} \mid T(\underline{x}) = t)\, P(T(\underline{x}) = t)$$

there is only one t for which the $P(\underline{X} = \underline{x} \mid T(\underline{X}) = t)$ is positive.

$$P(\underline{X} = \underline{x}) = \underbrace{P(\underline{X} = \underline{x} \mid T(\underline{X}) = t)}_{\text{independent of } \theta}\, P(T(\underline{X}) = t)$$

$$= h(\underline{X}) \qquad g(T(\underline{X}), \theta)$$

If $\underline{X} = (X_1, X_2, X_3)$

$$(X_1, X_2, X_3) = (x_1, x_2, x_3) \Rightarrow X_1 + X_2 + X_3 = x_1 + x_2 + x_2$$

<u>if part</u>:

in the if part we assume $f_\theta(\underline{x}) = g(T(\underline{x}), \theta)\, h(\underline{x})$ and we have to show that $\underline{X} \mid T(\underline{X})$ is independent of θ.

$$P(T(\underline{X}) = t) = \sum_{\underline{x} \in A_t} f_\theta(\underline{x})$$

where $A_t = \{(x_1, x_2, \ldots, x_n) : T(x_1, x_2, \ldots, x_n) = t\}$

$$= \sum_{\underline{x} \in A_t} g(T(\underline{x}), \theta)\, h(\underline{x})$$

$$= \sum_{\underline{x} \in A_t} g(t, \theta)\, h(\underline{x})$$

$$= g(t, \underline{\theta}) \sum_{\underline{x} \in A_t} h(\underline{x})$$

$$P(\underline{X} = \underline{x} \mid T(\underline{X}) = t) = \begin{cases} 0 & \text{if } t \neq T(\underline{x}) \\ \dfrac{P(\underline{X} = \underline{x})}{P(T(\underline{X}) = t)} & \text{if } t = T(\underline{x}) \end{cases}$$

$$\frac{P(\underline{X} = \underline{x})}{P(T(\underline{X}) = t)} = \frac{g(T(\underline{x}), \theta)\, h(\underline{x})}{g(t, \theta) \sum_{\underline{x} \in A_t} h(\underline{x})} = \frac{\cancel{g(t, \theta)}\, h(\underline{x})}{\cancel{g(t, \theta)} \sum_{\underline{x} \in A_t} h(\underline{x})}$$

free of θ.