Course Overview

C&B Chapters:

1: Probability Theory
2: Transformations / Expectations
3: Common Families of Distributions
4: Multiple Random Variables
5: Properties of a Random Sample
6: Principles of Data Reduction
7: Point Estimation
8: Hypothesis Testing
9: Interval Estimation
10: Asymptotic Evaluations

$\left.\begin{array}{l}1\text{–}5\end{array}\right\}$ 1st semester

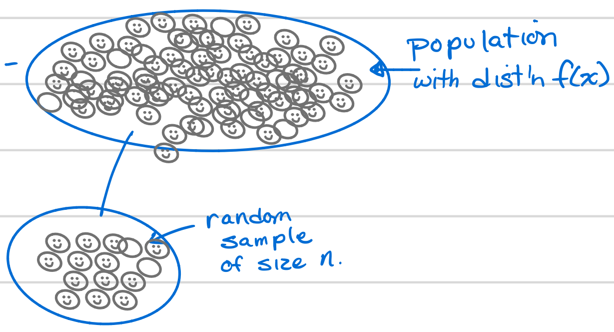
$\left.\begin{array}{l}6\text{–}10\end{array}\right\}$ 2nd semester

1st: Semester: Learning Tools
(info we'd like at our finger tips)

2nd semester: Applying Tools
(why we do what we do...)

§ 5.1 Basic Concepts of a Random Sample

- Random Sampling



Definition 5.1.1 The random variables $X_1, \ldots, X_n$ are called a *random sample of size n from the population $f(x)$* if $X_1, \ldots, X_n$ are mutually independent random variables and the marginal pdf or pmf of each X_i is the same function $f(x)$. Alternatively, $X_1, \ldots, X_n$ are called *independent and identically distributed random variables with pdf or pmf $f(x)$*. This is commonly abbreviated to iid random variables.

$X_1, \ldots, X_n$ random sample and are independent and identically distributed (iid Random Variables)

Since iid

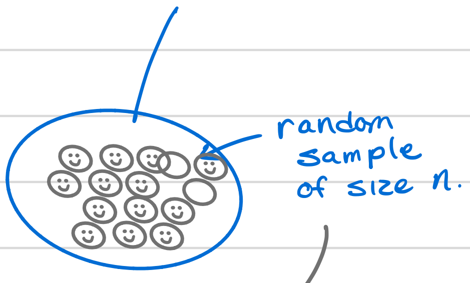
$$f(x_1, \ldots, x_n) = f(x_1) f(x_2) \cdot \ldots \cdot f(x_n) = \prod_{i=1}^{n} f(x_i)$$

If pdf is ft'n of parameter (θ)

$$f(x_1, \ldots, x_n \mid \theta) = \prod_{i=1}^{n} f(x_i \mid \theta)$$

Later (Chapter 7) we'll estimate θ from $X_1, \ldots, X_n$.

§ 5.2 Sums of RVs from a random sample



$Y = T(x_1, \ldots, x_n)$

Definition 5.2.1 Let $X_1, \ldots, X_n$ be a random sample of size n from a population and let $T(x_1, \ldots, x_n)$ be a real-valued or vector-valued function whose domain includes the sample space of $(X_1, \ldots, X_n)$. Then the random variable or random vector $Y = T(X_1, \ldots, X_n)$ is called a *statistic*. The probability distribution of a statistic Y is called the *sampling distribution of Y*.

Statistic : function of data not parameter, θ.

$\bar{X} = \Sigma X_i / n$

Definition 5.2.2 The *sample mean* is the arithmetic average of the values in a random sample. It is usually denoted by

$$\bar{X} = \frac{X_1 + \cdots + X_n}{n} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

$S^2 = \dfrac{\Sigma (X_i - \bar{X})^2}{(n-1)}$

Definition 5.2.3 The *sample variance* is the statistic defined by

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

The *sample standard deviation* is the statistic defined by $S = \sqrt{S^2}$.

least squares

$(n-1)S^2 = \Sigma x_i^2 - n\bar{x}^2$

Theorem 5.2.4 *Let $x_1, \ldots, x_n$ be any numbers and $\bar{x} = (x_1 + \cdots + x_n)/n$. Then*

a. $\min_a \sum_{i=1}^{n} (x_i - a)^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2$, ← least squares

b. $(n-1)s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$. ← simplifies calculation

$X_1, \ldots, X_n$ iid

$E[\Sigma g(X_i)] = n E[g(X_i)]$

$\mathrm{Var}[\Sigma g(X_i)] = n \mathrm{Var}[g(X_i)]$

Lemma 5.2.5 *Let $X_1, \ldots, X_n$ be a $\overbrace{\text{random sample}}^{\text{iid}}$ from a population and let $g(x)$ be a function such that $\mathrm{E}g(X_1)$ and $\mathrm{Var}\, g(X_1)$ exist. Then*

$$\mathrm{E}\left(\sum_{i=1}^{n} g(X_i)\right) = n\left(\mathrm{E}g(X_1)\right) \tag{5.2.1}$$

← true for identically distribute does not require independe

and

$$\mathrm{Var}\left(\sum_{i=1}^{n} g(X_i)\right) = n\left(\mathrm{Var}\, g(X_1)\right). \tag{5.2.2}$$

← true for iid

$X_1, \ldots, X_n$ iid

$E[X] = \mu$ $\mathrm{Var}[X] = \sigma^2$

- $E[\bar{X}] = \mu$
- $\mathrm{Var}[\bar{X}] = \sigma^2/n$
- $E[S^2] = \sigma^2$

Theorem 5.2.6 *Let $X_1, \ldots, X_n$ be a random sample from a population with mean μ and variance $\sigma^2 < \infty$. Then*

a. $\mathrm{E}\bar{X} = \mu$, ← unbiased estimator of μ.

b. $\mathrm{Var}\, \bar{X} = \dfrac{\sigma^2}{n}$,

c. $\mathrm{E}S^2 = \sigma^2$. ← unbiased estimator of σ^2.

mgf of $\bar{X}$

$M_{\bar{X}}(t) = [M_X(t/n)]^n$

Theorem 5.2.7 *Let $X_1, \ldots, X_n$ be a random sample from a population with mgf $M_X(t)$. Then the mgf of the sample mean is*

$$M_{\bar{X}}(t) = [M_X(t/n)]^n.$$

↑ useful if we recognize this mgf.

dist'n sums

- useful if mgf doesn't exist.

Theorem 5.2.9 *If X and Y are independent continuous random variables with pdfs $f_X(x)$ and $f_Y(y)$, then the pdf of $Z = X + Y$ is*

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(w) f_Y(z - w)\, dw. \tag{5.2.3}$$

exponential family

Statistic

$\sum_{j=1}^{n} t_1(X_j), \ldots, \sum_{j=1}^{n} t_k(X_j)$

is exponential family

Sampling from Exponential Family

Theorem 5.2.11 *Suppose $X_1, \ldots, X_n$ is a random sample from a pdf or pmf $f(x|\theta)$, where*

$$f(x|\theta) = h(x)c(\theta)\exp\left(\sum_{i=1}^{k} w_i(\theta)t_i(x)\right)$$

is a member of an exponential family. Define statistics $T_1, \ldots, T_k$ by (or $U_1, \ldots, U_n$)

$$U_i = T_i(X_1, \ldots, X_n) = \sum_{j=1}^{n} t_i(X_j), \quad i = 1, \ldots, k.$$

If the set $\{(w_1(\theta), w_2(\theta), \ldots, w_k(\theta)), \theta \in \Theta\}$ contains an open subset of $\Re^k$, then the distribution of $(T_1, \ldots, T_k)$ is an exponential family of the form

$$f_T(u_1, \ldots, u_k|\theta) = H(u_1, \ldots, u_k)[c(\theta)]^n \exp\left(\sum_{i=1}^{k} w_i(\theta)u_i\right). \tag{5.2.6}$$

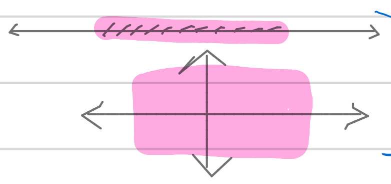
curved exponential

not open rectangle

$w_1(\theta) \ldots w_k(\theta)$ $\theta \in \Theta$ > open subset on $\mathbb{R}^k$

$\mathbb{R}^1$ $w_1(\theta)$ subset (not point) of $\mathbb{R}^1$

$\mathbb{R}^2$ $w_1(\theta)$ subset (not line) of $\mathbb{R}^2$



exclude curved exponential

$k > \dim(\Theta)$

§ 5.3 Sampling from Normal Dist'n

$X_1, \dots, X_n$ iid $N(\mu, \sigma^2)$

a) $\bar{X} \perp S^2$

b) $\bar{X} \sim N(\mu, \sigma^2/n)$

c) $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$

Theorem 5.3.1 *Let $X_1, \dots, X_n$ be a random sample from a $\mathrm{n}(\mu, \sigma^2)$ distribution, and let $\bar{X} = (1/n)\sum_{i=1}^n X_i$ and $S^2 = [1/(n-1)]\sum_{i=1}^n (X_i - \bar{X})^2$. Then*

a. *$\bar{X}$ and S^2 are independent random variables.*

b. *$\bar{X}$ has a $\mathrm{n}(\mu, \sigma^2/n)$ distribution,* proof example 5.2.8 (Notes pg 5).

c. *$(n-1)S^2/\sigma^2$ has a chi squared distribution with $n-1$ degrees of freedom.*

$Z \sim N(0,1)$

a) $Z^2 \sim \chi^2_1$

$X_i \sim \chi^2_{p_i}$

b) $\sum X_i \sim \chi^2_{\sum p_i}$

Lemma 5.3.2 (Facts about chi squared random variables) *We use the notation χ^2_p to denote a chi squared random variable with p degrees of freedom.*

a. *If Z is a $\mathrm{n}(0,1)$ random variable, then $Z^2 \sim \chi^2_1$; that is, the square of a standard normal random variable is a chi squared random variable.*

b. *If $X_1, \dots, X_n$ are independent and $X_i \sim \chi^2_{p_i}$, then $X_1 + \cdots + X_n \sim \chi^2_{p_1+\cdots+p_n}$; that is, independent chi squared variables add to a chi squared variable, and the degrees of freedom also add.*

$X_j \sim N(\mu_j, \sigma_j^2)$

$U_i = \sum a_{ij} X_j \quad i = 1 \dots k$

$V_r = \sum b_{rj} X_j \quad r = 1 \dots m$

$U_i \perp V_r \overset{\text{iff}}{\iff} \mathrm{Cov}(U_i, V_r) = 0$

$(U_1, \dots, U_k) \perp (V_1 \dots V_m)$

$\overset{\text{iff}}{\iff}$

$U_i \perp V_r \quad \forall i, r$

Lemma 5.3.3 *Let $X_j \sim \mathrm{n}(\mu_j, \sigma_j^2)$, $j = 1, \dots, n$, independent. For constants a_{ij} and b_{rj} ($j = 1, \dots, n$; $i = 1, \dots, k$; $r = 1, \dots, m$), where $k + m \le n$, define*

$$U_i = \sum_{j=1}^n a_{ij} X_j, \quad i = 1, \dots, k,$$

$$V_r = \sum_{j=1}^n b_{rj} X_j, \quad r = 1, \dots, m.$$

a. *The random variables U_i and V_r are independent if and only if $\mathrm{Cov}(U_i, V_r) = 0$. Furthermore, $\mathrm{Cov}(U_i, V_r) = \sum_{j=1}^n a_{ij} b_{rj} \sigma_j^2$.*

b. *The random vectors $(U_1, \dots, U_k)$ and $(V_1, \dots, V_m)$ are independent if and only if U_i is independent of V_r for all pairs i, r ($i = 1, \dots, k$; $r = 1, \dots, m$).*

$X_1, \dots, X_n$ iid $N(\mu, \sigma^2)$

$\sqrt{n}\dfrac{(\bar{X} - \mu)}{S} \sim t_{n-1}$

Definition 5.3.4 Let $X_1, \dots, X_n$ be a random sample from a $\mathrm{n}(\mu, \sigma^2)$ distribution. The quantity $(\bar{X} - \mu)/(S/\sqrt{n})$ has *Student's t distribution with $n-1$ degrees of freedom*. Equivalently, a random variable T has Student's t distribution with p degrees of freedom, and we write $T \sim t_p$ if it has pdf

$$f_T(t) = \frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)} \frac{1}{(p\pi)^{1/2}} \frac{1}{(1 + t^2/p)^{(p+1)/2}}, \quad -\infty < t < \infty. \tag{5.3.6}$$

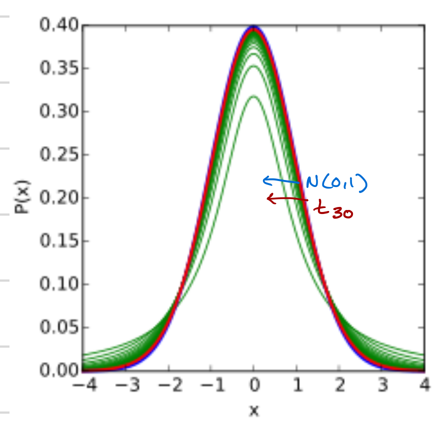
$X_1, \dots X_n$ iid $N(\mu_X, \sigma_X^2)$

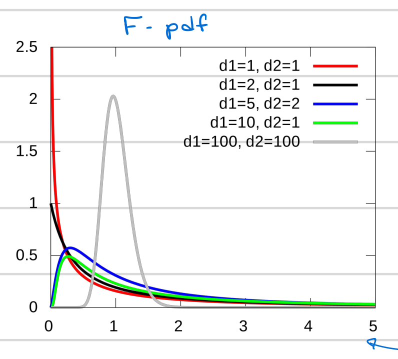
$Y_1, \dots, Y_m$ iid $N(\mu_Y, \sigma_Y^2)$

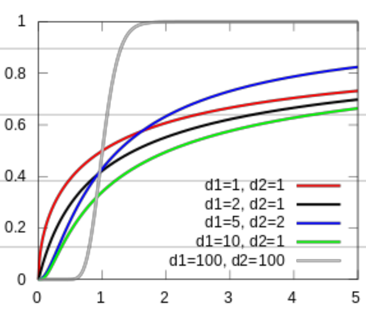
$\dfrac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \sim F_{n-1, m-1}$

Definition 5.3.6 Let $X_1, \dots, X_n$ be a random sample from a $\mathrm{n}(\mu_X, \sigma_X^2)$ population, and let $Y_1, \dots, Y_m$ be a random sample from an independent $\mathrm{n}(\mu_Y, \sigma_Y^2)$ population. The random variable $F = (S_X^2/\sigma_X^2)/(S_Y^2/\sigma_Y^2)$ has *Snedecor's F distribution with $n-1$ and $m-1$ degrees of freedom*. Equivalently, the random variable F has the F distribution with p and q degrees of freedom if it has pdf

$$f_F(x) = \frac{\Gamma\left(\frac{p+q}{2}\right)}{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{q}{2}\right)} \left(\frac{p}{q}\right)^{p/2} \frac{x^{(p/2)-1}}{[1 + (p/q)x]^{(p+q)/2}}, \quad 0 < x < \infty. \tag{5.3.9}$$







Order Stats

Definition 5.4.1 The *order statistics* of a random sample $X_1, \ldots, X_n$ are the sample values placed in ascending order. They are denoted by $X_{(1)}, \ldots, X_{(n)}$.

The order statistics are random variables that satisfy $X_{(1)} \le \cdots \le X_{(n)}$. In particular,

$$\begin{aligned} X_{(1)} &= \min_{1 \le i \le n} X_i, \\ X_{(2)} &= \text{second smallest } X_i, \\ &\vdots \\ X_{(n)} &= \max_{1 \le i \le n} X_i. \end{aligned}$$

percentiles

Definition 5.4.2 The notation $\{b\}$, when appearing in a subscript, is defined to be *the number b rounded to the nearest integer in the usual way.* More precisely, if i is an integer and $i - .5 \le b < i + .5$, then $\{b\} = i$.

Discrete pmf

Theorem 5.4.3 *Let $X_1, \ldots, X_n$ be a random sample from a discrete distribution with pmf $f_X(x_i) = p_i$, where $x_1 < x_2 < \cdots$ are the possible values of X in ascending order. Define*

$$\begin{aligned} P_0 &= 0 \\ P_1 &= p_1 \\ P_2 &= p_1 + p_2 \\ &\vdots \\ P_i &= p_1 + p_2 + \cdots + p_i \\ &\vdots \end{aligned}$$

Let $X_{(1)}, \ldots, X_{(n)}$ denote the order statistics from the sample. Then

$$P(X_{(j)} \le x_i) = \sum_{k=j}^{n} \binom{n}{k} P_i^k (1 - P_i)^{n-k} \tag{5.4.2}$$

j or more successes
$P(\text{success}) = P_i = P(X_\ell \le x_i)$
$\ell = 1, \ldots, n$ (iid)

and

$$P(X_{(j)} = x_i) = \sum_{k=j}^{n} \binom{n}{k} \left[P_i^k (1 - P_i)^{n-k} - P_{i-1}^k (1 - P_{i-1})^{n-k} \right]. \tag{5.4.3}$$

continuous pdf

Theorem 5.4.4 *Let $X_{(1)}, \ldots, X_{(n)}$ denote the order statistics of a random sample, $X_1, \ldots, X_n$, from a continuous population with cdf $F_X(x)$ and pdf $f_X(x)$. Then the pdf of $X_{(j)}$ is*

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x) [F_X(x)]^{j-1} [1 - F_X(x)]^{n-j}. \tag{5.4.4}$$

continuous

$f_{X_{(i)}, X_{(j)}}(u, v)$

Theorem 5.4.6 *Let $X_{(1)}, \ldots, X_{(n)}$ denote the order statistics of a random sample, $X_1, \ldots, X_n$, from a continuous population with cdf $F_X(x)$ and pdf $f_X(x)$. Then the joint pdf of $X_{(i)}$ and $X_{(j)}$, $1 \le i < j \le n$, is*

$$\begin{aligned} f_{X_{(i)}, X_{(j)}}(u, v) = {} & \frac{n!}{(i-1)!(j-1-i)!(n-j)!} f_X(u) f_X(v) [F_X(u)]^{i-1} \\ & \times [F_X(v) - F_X(u)]^{j-1-i} [1 - F_X(v)]^{n-j} \end{aligned} \tag{5.4.7}$$

for $-\infty < u < v < \infty$.

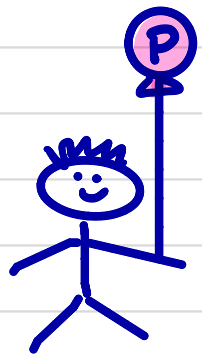
$f_{X_{(1)} \ldots X_{(n)}}(x_1, \ldots, x_n)$

$$f_{X_{(1)}, \ldots, X_{(n)}}(x_1, \ldots, x_n) = \begin{cases} n! f_X(x_1) \cdot \cdots \cdot f_X(x_n) & -\infty < x_1 < \cdots < x_n < \infty \\ 0 & \text{otherwise.} \end{cases}$$

$\hookrightarrow$ $n!$ permutations of $\underline{x}$ give same order stats

Where to next?

§5.5 Convergence Concepts



$\xrightarrow{P}$

Definition 5.5.1 A sequence of random variables, $X_1, X_2, \ldots$, *converges in probability* to a random variable X if, for every $\epsilon > 0$,

$$\lim_{n\to\infty} \underbrace{P(|X_n - X| \geq \epsilon)}_{\text{probability}} = 0 \quad \text{or, equivalently,} \quad \lim_{n\to\infty} P(|X_n - X| < \epsilon) = 1.$$

start here

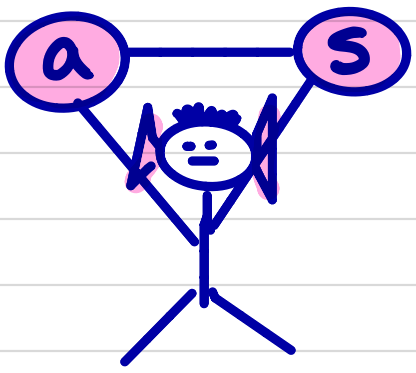
WLLN

$\bar{X}_n \xrightarrow{P} E[X]$

Theorem 5.5.2 (Weak Law of Large Numbers) *Let $X_1, X_2, \ldots$ be iid random variables with $\mathrm{E}X_i = \mu$ and $\mathrm{Var}\, X_i = \sigma^2 < \infty$. Define $\bar{X}_n = (1/n)\sum_{i=1}^n X_i$. Then, for every $\epsilon > 0$,*

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| < \epsilon) = 1;$$

that is, $\bar{X}_n$ converges in probability to μ.



$\xrightarrow{a.s.}$

Definition 5.5.6 A sequence of random variables, $X_1, X_2, \ldots$, *converges almost surely* to a random variable X if, for every $\epsilon > 0$,

$$P(\underbrace{\lim_{n\to\infty} |X_n - X| < \epsilon}_{\text{pointwise convergence}}) = 1.$$

start here

SLLN

$\bar{X}_n \xrightarrow{a.s.} E[X]$

Theorem 5.5.9 (Strong Law of Large Numbers) *Let $X_1, X_2, \ldots$ be iid random variables with $\mathrm{E}X_i = \mu$ and $\mathrm{Var}\, X_i = \sigma^2 < \infty$, and define $\bar{X}_n = (1/n)\sum_{i=1}^n X_i$. Then, for every $\epsilon > 0$,*

$$P(\lim_{n\to\infty} |\bar{X}_n - \mu| < \epsilon) = 1;$$

that is, $\bar{X}_n$ converges almost surely to μ.

$\xrightarrow{\mathcal{L}}$

Definition 5.5.10 A sequence of random variables, $X_1, X_2, \ldots$, *converges in distribution* to a random variable X if

convergence in Law

$$\lim_{n\to\infty} \underline{F_{X_n}(x)} = F_X(x)$$

start here (or mgf).

at all points x where $F_X(x)$ is continuous.

CLT

$f_{\bar{X}} \xrightarrow{\mathcal{L}} N(\mu, \sigma^2/n)$

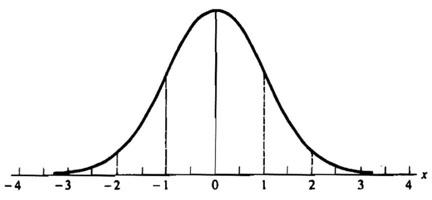


Figure 3.3.1. *Standard normal density*

Theorem 5.5.14 (Central Limit Theorem) *Let $X_1, X_2, \ldots$ be a sequence of iid random variables whose mgfs exist in a neighborhood of 0 (that is, $M_{X_i}(t)$ exists for $|t| < h$, for some positive h). Let $\mathrm{E}X_i = \mu$ and $\mathrm{Var}\, X_i = \sigma^2 > 0$. (Both μ and σ^2 are finite since the mgf exists.) Define $\bar{X}_n = (1/n)\sum_{i=1}^n X_i$. Let $G_n(x)$ denote the cdf of $\sqrt{n}(\bar{X}_n - \mu)/\sigma$. Then, for any x, $-\infty < x < \infty$,*

$$\lim_{n\to\infty} G_n(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy;$$

that is, $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ has a limiting standard normal distribution.

$$\therefore \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \xrightarrow{\mathcal{L}} N(0,1)$$

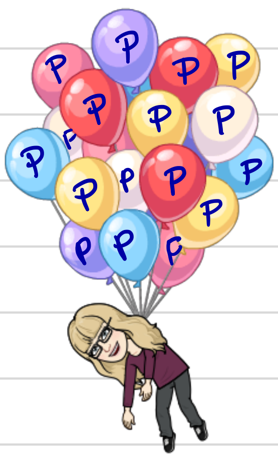
proof: homework-2

Theorem 5.5.15 (Stronger form of the Central Limit Theorem) *Let $X_1, X_2, \ldots$ be a sequence of iid random variables with $\mathrm{E}X_i = \mu$ and $0 < \mathrm{Var}\, X_i = \sigma^2 < \infty$. Define $\bar{X}_n = (1/n)\sum_{i=1}^n X_i$. Let $G_n(x)$ denote the cdf of $\sqrt{n}(\bar{X}_n - \mu)/\sigma$. Then, for any x, $-\infty < x < \infty$,*

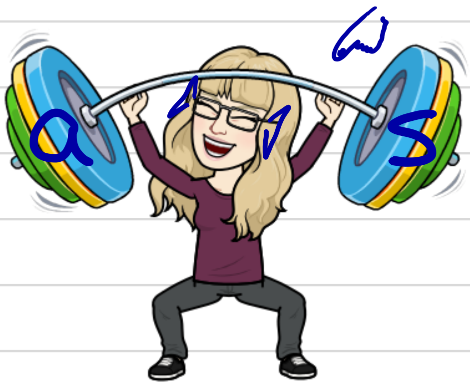
$$\lim_{n\to\infty} G_n(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy;$$

that is, $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ has a limiting standard normal distribution.

proof uses characteristic ft'n $E[e^{itX}]$ $(i^2 = -1)$



$\xrightarrow{P}$



$\xrightarrow{a.s.}$

$\xrightarrow{\mathcal{L}}$

$\xrightarrow{P}$ implies $\xrightarrow{\mathcal{L}}$

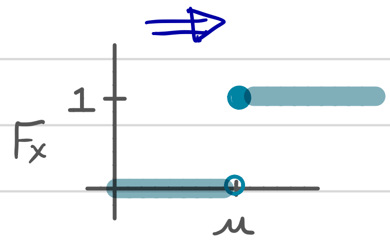
Recall

$\xrightarrow{a.s.} \;\; \cancel{\Rightarrow} \;\; \xrightarrow{P}$

Theorem 5.5.12 *If the sequence of random variables, $X_1, X_2, \ldots$, converges in probability to a random variable X, the sequence also converges in distribution to X.*

$X_n \xrightarrow{P} \mu$

$\Rightarrow$



Theorem 5.5.13 *The sequence of random variables, $X_1, X_2, \ldots$, converges in probability to a constant μ if and only if the sequence also converges in distribution to μ. That is, the statement*

$$P(|X_n - \mu| > \varepsilon) \to 0 \text{ for every } \varepsilon > 0$$

is equivalent to

$$P(X_n \le x) \to \begin{cases} 0 & \text{if } x < \mu \\ 1 & \text{if } x > \mu. \end{cases}$$

if: $X_n \xrightarrow{\mathcal{L}} X \quad Y_n \xrightarrow{P} a$

$Y_n X_n \xrightarrow{\mathcal{L}} aX$

$X_n + Y_n \xrightarrow{\mathcal{L}} X + a$

Theorem 5.5.17 (Slutsky's Theorem) *If $X_n \to X$ in distribution and $Y_n \to a$, a constant, in probability, then*

a. *$Y_n X_n \to aX$ in distribution.*

b. *$X_n + Y_n \to X + a$ in distribution.*

By Slutsky's

$$\frac{\sqrt{n}(\bar{X} - \mu)}{S} \xrightarrow{\mathcal{L}} N(0,1)$$

if $\bar{X} \xrightarrow{\mathcal{L}} N(\mu, \sigma^2/n)$ CLT

$S^2 \xrightarrow{P} \sigma^2$ Cheby's

Example: (Normal approx with estimated variance)

Suppose $\dfrac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \xrightarrow{\mathcal{L}} N(0,1)$

But σ is unknown. Last lecture showed $S_n^2 \xrightarrow{P} \sigma^2$ (using Cheby's)

$\Rightarrow S_n \longrightarrow \sigma$

$\Rightarrow \sigma/S_n \longrightarrow 1$

By Slutsky's Thm: $\dfrac{\sqrt{n}(\bar{X} - \mu)}{S_n} = \underbrace{\left(\dfrac{\sigma}{S_n}\right)}_{\xrightarrow{P} 1} \underbrace{\left(\dfrac{\sqrt{n}(\bar{X} - \mu)}{\sigma}\right)}_{\xrightarrow{\mathcal{L}} N(0,1)} \xrightarrow{\mathcal{L}} N(0,1)$

Taylor's expansion of $g(x)$ about a (order r)

$r=1$

$T_1(x) = g(a) + g'(a)(x-a)$

$r=2$

$T_2(x) = g(a) + g'(a)(x-a) + g''(a)(x-a)^2/2$

Definition 5.5.20 If a function $g(x)$ has derivatives of order r, that is, $g^{(r)}(x) = \frac{d^r}{dx^r} g(x)$ exists, then for any constant a, the *Taylor polynomial of order r about a* is

$$T_r(x) = \sum_{i=0}^{r} \frac{g^{(i)}(a)}{i!}(x-a)^i.$$

Taylor's major theorem, which we will not prove here, is that the *remainder* from the approximation, $g(x) - T_r(x)$, always tends to 0 faster than the highest-order explicit term.

Taylor's Thm:

$$\lim_{x \to a} \frac{g(x) - T_r(x)}{(x-a)^r} = 0$$

Theorem 5.5.21 (Taylor) *If $g^{(r)}(a) = \frac{d^r}{dx^r} g(x)\big|_{x=a}$ exists, then*

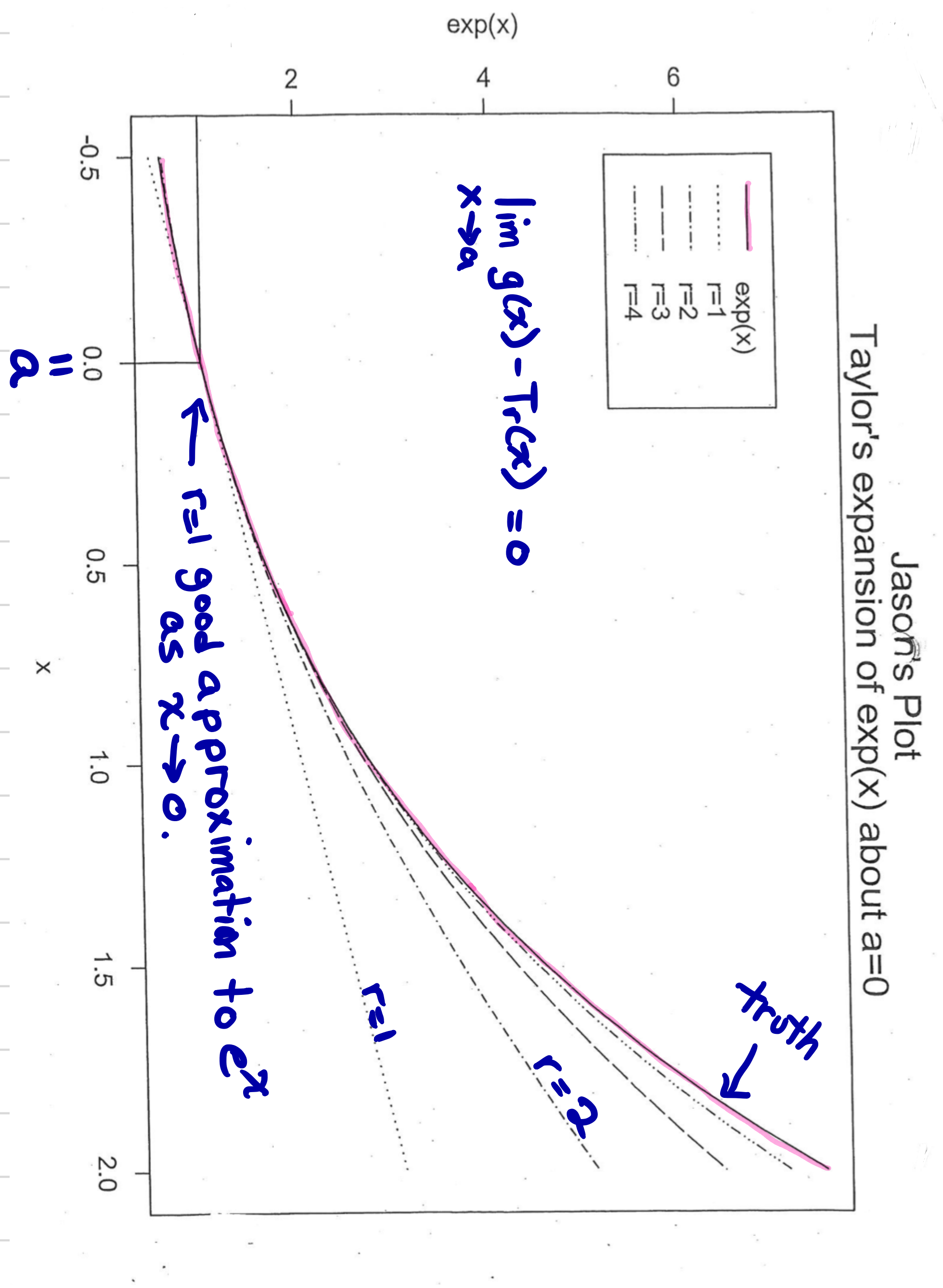
$$\lim_{x \to a} \frac{g(x) - T_r(x)}{(x-a)^r} = 0.$$

In general, we will not be concerned with the explicit form of the remainder. Since we are interested in approximations, we are just going to ignore the remainder. There are, however, many explicit forms, one useful one being

$$g(x) - T_r(x) = \int_a^x \frac{g^{(r+1)}(t)}{r!}(x-t)^r dt.$$

Taylor's expansion
$r = 1, 2, 3, 4$

of $g(x) = e^x$
$a = 0$



§ 5.5.4 The Delta Method

$X_1 \dots X_n$ iid Bernoulli (p)

estimate p by $\bar{X}$

estimate $\frac{p}{1-p}$ by $\frac{\bar{X}}{1-\bar{X}}$

how estimate

$Var\left(\frac{\bar{X}}{1-\bar{X}}\right)$?

or dist'n $\frac{\bar{X}}{1-\bar{X}}$?

Suppose $X_1, \dots X_n$ iid Bernoulli(p), we may be interested in estimating p.

may also be interested in estimating odds: $\left(\frac{p}{1-p}\right)$

or biostatisticians (or epidemiologists) may be interested in estimating odds ratio for two treatments: $\left(\left(\frac{p}{(1-p)}\right) / \left(\frac{r}{(1-r)}\right)\right)$

- assume estimate p by $\frac{\Sigma X_i}{n} = \frac{\#\text{ success}}{\#\text{ trials}} = \hat{p}$

- then could estimate $\frac{p}{1-p}$ by $\frac{\hat{p}}{1-\hat{p}} = \frac{\Sigma X_i / n}{1 - \Sigma X_i/n}$

- How estimate $Var\left(\frac{\hat{p}}{1-\hat{p}}\right)$ or it's sampling dist'n ?

"Intuition abandons us, and exact calculation is relatively hopeless, so we have to rely on an approximation."

Show that can approximate

mean

$E[g(X)] \approx g(\mu)$

variance

$Var[g(X)] \approx (g'(\mu))^2 \, Var[X]$

Example (Approximate mean + variance)

- assume X is R.V. $E[X] = \mu \neq 0$ [here $n=1$]
- wish to estimate $g(\mu)$
- First order Taylor's:

$$g(X) = g(\mu) + g'(\mu)(X-\mu)$$

$$E[g(X)] \approx g(\mu)$$

$$Var[g(X)] \approx (g'(\mu))^2 \, Var[X]$$

Suppose: $g(\mu) = 1/\mu$

estimate $1/\mu$ by $1/X$ since $E[1/X] \approx 1/\mu$

$$Var[1/X] \approx (1/\mu)^4 \, Var(X)$$

hmm... again f'th of unknown μ.

- also what if don't know $Var(X)$?

Approximate dist'n

if $\sqrt{n}(Y_n - \theta) \xrightarrow{d} N(0, \sigma^2)$

then approximate dist'n

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

Theorem 5.5.24 (Delta Method) *Let Y_n be a sequence of random variables that satisfies $\sqrt{n}(Y_n - \theta) \to \mathrm{n}(0, \sigma^2)$ in distribution. For a given function g and a specific value of θ, suppose that $g'(\theta)$ exists and is not 0. Then*

$$\sqrt{n}[g(Y_n) - g(\theta)] \to \mathrm{n}(0, \sigma^2[g'(\theta)]^2) \text{ in distribution.} \tag{5.5.10}$$

Proof (Taylor's + Slutsky's).

approximate
mean (Taylor's)
variance (Taylor's)
dist'n (Delta)

Additional approximation since σ^2 unknown

$\xrightarrow{P}$ $\xrightarrow{d}$

Slutsky's.

Continuing above example (now have random sample, rather than $n=1$).

$E[\bar{X}] = \mu$ assuming finite $Var(X_1) = \sigma^2$ $\quad (Var(\bar{X}) = \sigma^2/n)$

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

by Delta: $\sqrt{n}\left(\frac{1}{\bar{X}} - \frac{1}{\mu}\right) \xrightarrow{d} N\left(0, \left(\frac{1}{\mu}\right)^4 \sigma^2\right)$

or $Var\left(\frac{1}{\bar{X}}\right) \approx \left(\frac{1}{\mu}\right)^4 \sigma^2$

know $\bar{X} \xrightarrow{P} \mu$ WLLN, $S^2 \to \sigma^2$ Cheby's — Lecture 3

$\Rightarrow \left(\frac{1}{\bar{X}}\right)^4 \xrightarrow{P} \left(\frac{1}{\mu}\right)^4$

$$\left(\frac{1}{\bar{X}}\right)^4 S^2 \xrightarrow{P} \left(\frac{1}{\mu}\right)^4 \sigma^2$$

Theorem 5.5.4 *Suppose that $X_1, X_2, \ldots$ converges in probability to a random variable X and that h is a continuous function. Then $h(X_1), h(X_2), \ldots$ converges in probability to $h(X)$.*

Summary

$\sqrt{n}\left(\frac{1}{\bar{X}} - \frac{1}{\mu}\right) \xrightarrow{d} N\left(0, \left(\frac{1}{\mu}\right)^4 \sigma^2\right)$

$\left(\frac{1}{\bar{X}}\right)^4 S^2 \xrightarrow{P} \left(\frac{1}{\mu}\right)^4 \sigma^2$

Convergence in law $\xrightarrow{d}$, Convergence in prob $\xrightarrow{P}$ — Slutsky's

Theorem 5.5.17 (Slutsky's Theorem) *If $X_n \to X$ in distribution and $Y_n \to a$, a constant, in probability, then*

a. *$Y_n X_n \to aX$ in distribution.*

b. *$X_n + Y_n \to X + a$ in distribution.*

$$\frac{\sqrt{n}\left(\frac{1}{\bar{X}} - \frac{1}{\mu}\right)}{\left(\frac{1}{\bar{X}}\right)^2 S} = \underbrace{\frac{\sqrt{n}\left(\frac{1}{\bar{X}} - \frac{1}{\mu}\right)}{\left(\frac{1}{\mu}\right)^2 \sigma}}_{\xrightarrow{d} N(0,1)} * \underbrace{\frac{\left(\frac{1}{\mu}\right)^2 \sigma}{\left(\frac{1}{\bar{X}}\right)^2 S}}_{\xrightarrow{P} 1} \xrightarrow{d} N(0,1)$$

Extensions of Delta Method

What if $g'(\theta)=0$ but $g''(\theta)\neq 0$

how approximate variance?

If $g'(\theta)=0$ $\longrightarrow$

Theorem 5.5.26 (Second-order Delta Method) *Let Y_n be a sequence of random variables that satisfies $\sqrt{n}(Y_n-\theta)\to \mathrm{n}(0,\sigma^2)$ in distribution. For a given function g and a specific value of θ, suppose that $g'(\theta)=0$ and $g''(\theta)$ exists and is not 0. Then*

$$n[g(Y_n)-g(\theta)]\to \sigma^2\frac{g''(\theta)}{2}\chi_1^2 \text{ in distribution.} \tag{5.5.13}$$

θ is vector of length p

$(\mu_1,\mu_2,\dots\mu_p)$

Note $g(\bar{X}_1,\dots,\bar{X}_p)$ scalar

$E[X_{ij}]=\mu_i$ $\quad i=1,\dots p$

$E[\bar{X}_i]=\mu_i$

$\sqrt{n}\,[g(\bar{X}_1,\dots,\bar{X}_p)-g(\mu_1,\dots\mu_p)] \xrightarrow{\mathcal{L}} N(0,\tau^2)$ scalar

If $\underset{\sim}{\theta}$ is vector

vector of length p — n patients, each with p observations.

Theorem 5.5.28 (Multivariate Delta Method) *Let $\mathbf{X}_1,\dots,\mathbf{X}_n$ be a random sample with $\mathrm{E}(X_{ij})=\mu_i$ and $\mathrm{Cov}(X_{ik},X_{jk})=\sigma_{ij}$. For a given function g with continuous first partial derivatives and a specific value of $\boldsymbol{\mu}=(\mu_1,\dots,\mu_p)$ for which $\tau^2=\Sigma\Sigma\sigma_{ij}\frac{\partial g(\mu)}{\partial\mu_i}\cdot\frac{\partial g(\mu)}{\partial\mu_j}>0$,*

$$\sqrt{n}[g(\bar{X}_1,\dots,\underset{s=p\,?}{\bar{X}_s})-g(\mu_1,\dots,\mu_p)]\to \mathrm{n}(0,\tau^2) \text{ in distribution}.$$

If $\sigma_{ij}=Cov(X_i,X_j)$

$\sigma_i^2=Var(X_i)=Cov(X_i,X_i)$

$\rightarrow$ Define symmetric Variance / Covariance matrix

Var/Cov matrix symmetric $\sigma_{ij}=\sigma_{ji}$

$$\tau^2=\underbrace{\underset{(1\times p)}{\begin{bmatrix} g'(\theta_1)\cdots g'(\theta_p)\end{bmatrix}}\;\underset{(p\times p)}{\begin{bmatrix}\sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p}\\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2p}\\ \vdots & & \ddots & \\ \sigma_{1p} & & & \sigma_p^2\end{bmatrix}}\;\underset{(p\times 1)}{\begin{bmatrix} g_1'(\theta)\\ \vdots\\ g_p'(\theta)\end{bmatrix}}}_{1\times 1}$$

Note: This version of multivariate delta is ft'ns of $\bar{X}_1,\dots,\bar{X}_s$.
n is sample size p is # parameters (# θs).

assumes $(\bar{X}_1,\dots\bar{X}_p)\to MVN\left[\begin{pmatrix}\mu_1\\ \vdots\\ \mu_p\end{pmatrix},\ \underset{\text{var/cov matrix}}{\Sigma}\right]$ by CLT...

Sufficiency Summary / Tools

SUFFICIENCY PRINCIPLE: If $T(\mathbf{X})$ is a sufficient statistic for θ, then any inference about θ should depend on the sample $\mathbf{X}$ only through the value $T(\mathbf{X})$. That is, if $\mathbf{x}$ and $\mathbf{y}$ are two sample points such that $T(\mathbf{x}) = T(\mathbf{y})$, then the inference about θ should be the same whether $\mathbf{X} = \mathbf{x}$ or $\mathbf{X} = \mathbf{y}$ is observed.

	x	y	
HHH	HHT	HTH	THH
TTH	THT	HTT	TTT

Sufficiency Principle (C&B):

*If T(**X**) is a sufficient statistic for θ then any inference about θ should depend on the sample **X** only through the value T(**X**).*

*That is, if **x** and **y** are two sample points such that T(**x**)=T(**y**), then the inference about θ should be the same whether **X**=**x** or **Y**=**y** is observed.*

Definition 6.2.1 A statistic $T(\mathbf{X})$ is a *sufficient statistic for* θ if the conditional distribution of the sample $\mathbf{X}$ given the value of $T(\mathbf{X})$ does not depend on θ.

That is: $P_\theta(\underline{X}=\underline{x} \mid T(\underline{X})=T(\underline{x}))$

$= P(\underline{X}=\underline{x} \mid T(\underline{X})=T(\underline{x})) \quad (\perp \theta)$

*A statistic T(**X**) is a sufficient statistic for θ if the conditional dist'n of the sample **X** given the value T(**X**) does not depend on θ.*

Theorem 6.2.2 *If $p(\mathbf{x}|\theta)$ is the joint pdf or pmf of $\mathbf{X}$ and $q(t|\theta)$ is the pdf or pmf of $T(\mathbf{X})$, then $T(\mathbf{X})$ is a sufficient statistic for θ if, for every $\mathbf{x}$ in the sample space, the ratio $p(\mathbf{x}|\theta)/q(T(\mathbf{x})|\theta)$ is constant as a function of θ.*

Coin flip example

- $p(\mathbf{x}|\theta) = p^{\Sigma x}(1-p)^{3-\Sigma x}$
- $q(\mathbf{t}|\theta) = \binom{3}{\Sigma x} p^{\Sigma x}(1-p)^{3-\Sigma x}$
- $p(\mathbf{x}|\theta) / q(\mathbf{t}|\theta) = 1/\binom{3}{\Sigma x}$

*If **X**~ p(**x**|θ) and T(**X**)~ q(**t**|θ)*

*Then for every **x** in the sample space:*

*T(**X**) sufficient* $\rightarrow$ *p(**x**|θ) / q(T(**X**)|θ) constant wrt θ*

*T(**X**) sufficient* $\leftarrow$ *p(**x**|θ) / q(T(**X**)|θ) constant wrt θ*

Theorem 6.2.6 (Factorization Theorem) *Let $f(\mathbf{x}|\theta)$ denote the joint pdf or pmf of a sample $\mathbf{X}$. A statistic $T(\mathbf{X})$ is a sufficient statistic for θ if and only if there exist functions $g(t|\theta)$ and $h(\mathbf{x})$ such that, for all sample points $\mathbf{x}$ and all parameter points θ,*

$$f(\mathbf{x}|\theta) = g(T(\mathbf{x})|\theta)h(\mathbf{x}). \tag{6.2.3}$$

Theorem 6.2.6 (Factorization Theorem) *Let $f(\mathbf{x}|\theta)$ denote the joint pdf or pmf of a sample $\mathbf{X}$. A statistic $T(\mathbf{X})$ is a sufficient statistic for θ if and only if there exist functions $g(t|\theta)$ and $h(\mathbf{x})$ such that, for all sample points $\mathbf{x}$ and all parameter points θ,*

$$f(\mathbf{x}|\theta) = g(T(\mathbf{x})|\theta)h(\mathbf{x}). \tag{6.2.3}$$

$f(\mathbf{x}|\theta) = p^{\Sigma x}(1-p)^{3-\Sigma x}$

$g(T(\mathbf{x})|\theta) = p^{\Sigma x}(1-p)^{3-\Sigma x}$

$h(x) = 1$

$T(\mathbf{x}) = \Sigma x$ = sufficient

Sample $\mathbf{X}\sim f(\mathbf{x}|\theta)$, i.e. $f(\mathbf{x}|\theta)$ is the joint dist'n

$T(\mathbf{X})$ sufficient $\rightarrow$ $f(\mathbf{x}|\theta)=g(T(\mathbf{x})|\theta)h(\mathbf{x})$

$T(\mathbf{X})$ sufficient $\leftarrow$ $f(\mathbf{x}|\theta)=g(T(\mathbf{x})|\theta)h(\mathbf{x})$

Functions $g(t|\theta)$ and $h(\mathbf{x})$ exist for all $\mathbf{x}$ and all parameter points θ.

HHH	HHT	HTH	THH
TTH	THT	HTT	TTT

Theorem 6.2.10 *Let $X_1, \ldots, X_n$ be iid observations from a pdf or pmf $f(x|\boldsymbol{\theta})$ that belongs to an exponential family given by*

$$f(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta})\exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta})t_i(x)\right),$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_d)$, $d \le k$. Then

$$T(\mathbf{X}) = \left(\sum_{j=1}^{n} t_1(X_j), \ldots, \sum_{j=1}^{n} t_k(X_j)\right)$$

is a sufficient statistic for $\boldsymbol{\theta}$.

$f(x|\theta) = p^x(1-p)^{1-x}$ (one x, Bernoulli dist'n)

$= (p/(1-p))^x (1-p)^1$

$= (1-p)\exp\{x \log(p/(1-p))\}$

$t(x)=x$, $h(x)=1$,

$c(\theta)=(1-p)$, $w(\theta)=\log(p/(1-p))$

$X_1 \ldots X_n \sim f(x|\theta)$, iid

$f(x|\theta)= h(x)\, c(\theta) \exp(w(\theta)\, t(x))$ *(note: for an individual X)*

Then

$T(X) = \sum_{j=1}^{n} t(x_j)$

is a sufficient statistic for θ.

*Assuming exponential family, important information for qualifying exams.

Not sufficient (lose sufficiency when separate same)

HHH	HHT	HTH	THH
HTT	THT	TTH	TTT

HHH	HHT	HTH	THH
HTT	THT	TTH	TTT

HHH	HHT	HTH	THH
HTT	THT	TTH	TTT

HHH	HHT	HTH	THH
HTT	THT	TTH	TTT

↑ not sufficient
not complete

Definition 6.2.11 A sufficient statistic $T(\mathbf{X})$ is called a *minimal sufficient statistic* if, for any other sufficient statistic $T'(\mathbf{X})$, $T(\mathbf{x})$ is a function of $T'(\mathbf{x})$.

Sufficient	**X** = full data vector
Sufficient	U(**X**) a reduction of the data
Minimal Sufficient Further reduction, lose sufficiency	T(**X**) = h(U(**X**)) a further reduction

- Def'n (C&B)
  *A sufficient statistics T(**X**) is called a minimal sufficient statistic if, for any other sufficient statistic T'(**X**), T(**x**) is a function of T'(**x**).*

Theorem 6.2.13 *Let $f(\mathbf{x}|\theta)$ be the pmf or pdf of a sample $\mathbf{X}$. Suppose there exists a function $T(\mathbf{x})$ such that, for every two sample points $\mathbf{x}$ and $\mathbf{y}$, the ratio $f(\mathbf{x}|\theta)/f(\mathbf{y}|\theta)$ is constant as a function of θ if and only if $T(\mathbf{x}) = T(\mathbf{y})$. Then $T(\mathbf{X})$ is a minimal sufficient statistic for θ.*

$f(\mathbf{x}|\theta) = p^{\Sigma x}(1-p)^{3-\Sigma x}$

$f(\mathbf{y}|\theta) = p^{\Sigma y}(1-p)^{3-\Sigma y}$

$f(\mathbf{x}|\theta) / f(\mathbf{y}|\theta) = (p/(1-p))^{\Sigma x - \Sigma y}$

$f(\mathbf{x}|\theta) / f(\mathbf{y}|\theta)$ constant wrt θ if $\Sigma x = \Sigma y$

$T(\mathbf{x}) = \Sigma x$ = minimal sufficient

*Sample **X**~ f(**x**| θ), i.e. f(**x**| θ) is the joint dist'n*

*T(**X**) minimal sufficient → f(**x**| θ) / f(**y**| θ) constant wrt θ if T(**x**) = T(**y**).*

*T(**X**) minimal sufficient ← f(**x**| θ) / f(**y**| θ) constant wrt θ if T(**x**) = T(**y**).*

HHH	HHT	HTH	THH
TTH	THT	HTT	TTT

Definition 6.2.16 A statistic $S(\mathbf{X})$ whose distribution does not depend on the parameter θ is called an *ancillary statistic*.

*A statistic S(**X**) whose dist'n doesn't depend on θ is called ancillary. (S is for ancillary)*

We may have a minimal sufficient statistic,
T' =(n, Σx) with dimension(T') > dimension(θ).

Coin flip: T' =(n, Σx), θ = Pr(H) = p

We can think of Σx as 'conditionally sufficient' because it is used as a sufficient statistic in inference conditional on N=n.

Definition 6.2.21 Let $f(t|\theta)$ be a family of pdfs or pmfs for a statistic $T(\mathbf{X})$. The family of probability distributions is called *complete* if $\mathrm{E}_\theta g(T) = 0$ for all θ implies $P_\theta(g(T) = 0) = 1$ for all θ. Equivalently, $T(\mathbf{X})$ is called a *complete statistic*.

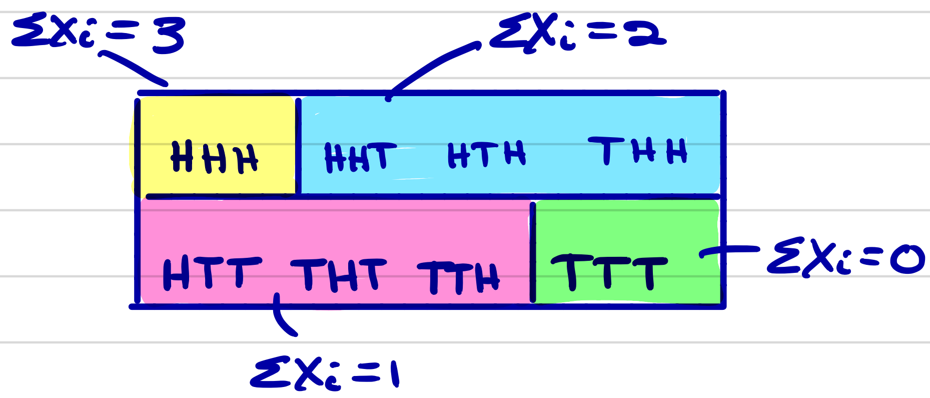
Let $f(t|\theta)$ be a family of pdfs or pmfs for a statistic $T(\mathbf{X})$. The family of probability dist'ns is called complete if

$E_\theta[g(T)] = 0$ for all $\theta \rightarrow P_\theta(g(T)=0)=1$ for all θ.

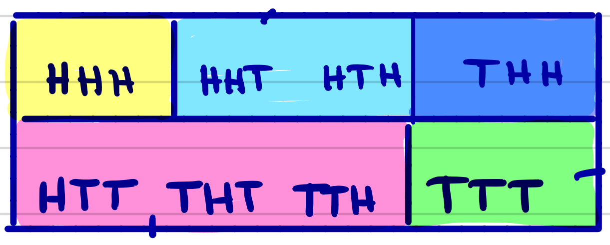
Equivalently, $T(\mathbf{X})$ is called a complete statistic.

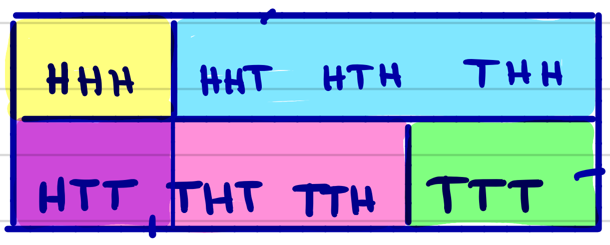
Heuristic Interpretation:

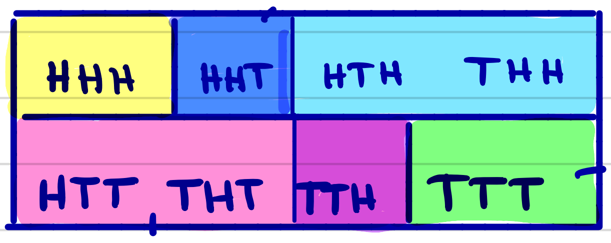
Flipping coin:
Partition sample space based on # heads is complete.

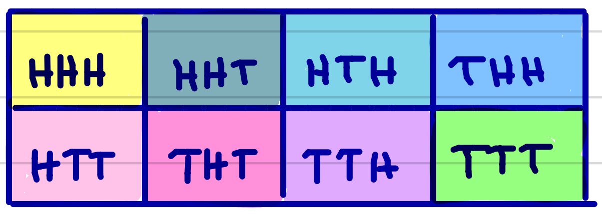


Not Complete:









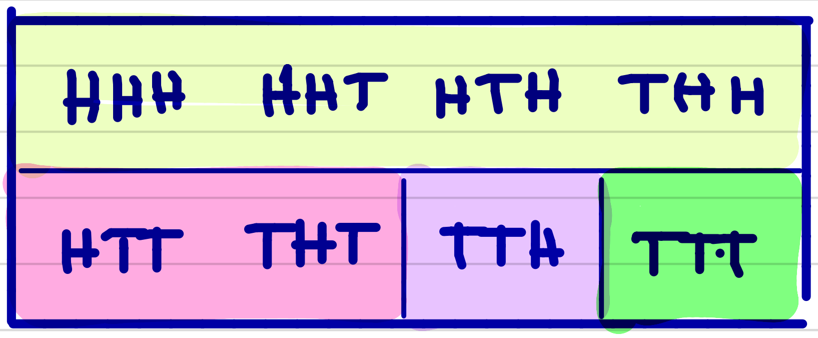
If separate data partitions that have same $T(\underline{x})$ (sufficient statistic)
$\rightarrow$ Not Complete. But sufficient

Coin Flip Example: Completeness

- *$E_\theta[g(T)] = 0$ for all $\theta \rightarrow P_\theta(g(T)=0)=1$ for all θ.*
- T(X) ~ binomial(3,p) (assuming n=3 known)
- $q(\mathbf{t}|\theta) = \binom{3}{\Sigma x} p^{\Sigma x}(1-p)^{3-\Sigma x}$
- E[g(T)] = $g(0)*1*p^0(1-p)^{3-0} + g(1)*3*p^1(1-p)^{3-1} + g(2)*3*p^2(1-p)^{3-2} + g(3)*1*p^3(1-p)^{3-3}$

  $= (1-p)^3\,[g(0)*1*(p/(1-p))^0 + g(1)*3*(p/(1-p))^1 + g(2)*3*(p/(1-p))^2 + g(3)*1*(p/(1-p))^3]$

  $=?\ 0 \rightarrow g(0)=g(1)=g(2)=g(3) = 0$



$\uparrow$ not sufficient
not complete

Relationship between Sufficiency and Completeness

T(**X**) = h(U(**X**))
U(**X**): reduction of the data
T(**X**) = h(U(**X**)):further reduction

T sufficient → U sufficient
T complete ← U complete

Sufficient	**X** = full data vector	
Sufficient	U(**X**) a reduction of the data	At some point lose completeness
Minimal Sufficient Further reduction, lose sufficiency	T(**X**) = h(U(**X**)) a further reduction	Complete ?
	further reduction	Complete
	Constant, most extreme reduction	Complete

combine different lose sufficiency

Separate same (too many partitions)

lose Completeness

Relationship between Sufficiency and Completeness

T(**X**) = h(U(**X**))
U(**X**): reduction of the data
T(**X**) = h(U(**X**)):further reduction

T sufficient → U sufficient
T complete ← U complete

Sufficient	HHH TTH	HHT THT	HTH HTT	THH TTT	Finest partition (8) (X_1, X_2, X_3)
Sufficient	HHH TTH	HHT THT	HTH HTT	THH TTT	U(**X**) = Reduction of data (6 partitions)
Minimal Sufficient / Complete	HHH TTH	HHT THT	HTH HTT	THH TTT	T(**X**) Grouped by ΣX_i # Heads (4 partitions)
Complete	HHH TTH	HHT THT	HTH HTT	THH TTT	T_1(**X**) = further reduction. (3 partitions)
Complete	HHH TTH	HHT THT	HTH HTT	THH TTT	Coarsest partition (1) T_2(**X**)=Constant for all **X**

Collecting data

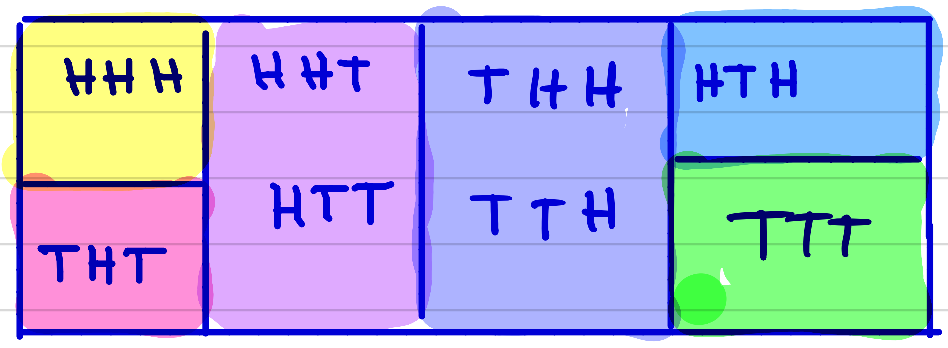
Anal Anna	Dyslexic Dan	Careless Kim	Nirvana Nell
HHT	HHH	flipped 3 times	# heads
HTH	HHT or HTT	Yes/No	3
HHH	TTT		2
⋮	HTH		3
	TTH or THH		1
	⋮		0
			⋮

collect all infomation

Anal Anna
Sufficient Not Complete

HHH	HHT	HTH	THH
TTH	THT	HTT	TTT

Dsylexic Dan
Not sufficient, Not Complete



Careless Kim
Complete, Not Sufficient

1 HHH HHT HTH THH
TTH THT HTT TTT

Nirvana Nell

HHH	HHT	HTH	THH
TTH	THT	HTT	TTT

Why do we care about completeness? ancillary? minimal sufficient?

redundant

Theorem 6.2.24 (Basu's Theorem) *If $T(\mathbf{X})$ is a complete and minimal sufficient statistic, then $T(\mathbf{X})$ is independent of every ancillary statistic.*

Complete and sufficient $\Longrightarrow$ minimal sufficient

$\not\Longleftarrow$

Most problems covered by exponential family

Theorem 6.2.25 (Complete statistics in the exponential family) *Let $X_1, \ldots, X_n$ be iid observations from an exponential family with pdf or pmf of the form*

$$f(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp\left(\sum_{j=1}^{k} w(\theta_j)t_j(x)\right), \tag{6.2.7}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$. Then the statistic

$$T(\mathbf{X}) = \left(\sum_{i=1}^{n} t_1(X_i), \sum_{i=1}^{n} t_2(X_i), \ldots, \sum_{i=1}^{n} t_k(X_i)\right)$$

is complete as long as the parameter space Θ contains an open set in $\Re^k$.

sufficient (C&B errata)

Theorem 6.2.28 *If a minimal sufficient statistic exists, then any complete sufficient statistic is also a minimal sufficient statistic.*