Math 4501 - Probability and Statistics II

5.6 - Central Limit Theorem (review)

from 3501

Central Limit Theorem

$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} x_i$

Theorem

Let $\bar{X}$ be the mean of a $\overbrace{\text{random sample}}^{\text{iid}}$ $X_1, X_2, \ldots, X_n$ of size n from a distribution with a finite mean μ and a finite positive variance σ^2.

Then the distribution of

$$W = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \overset{\times n}{=} \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma}$$

approaches a standard normal distribution $N(0, 1)$ in the limit as $n \to \infty$.

increasing sample size

For large n, the distribution of $\bar{X}$ is close to $N(\mu, \frac{\sigma^2}{n})$ and so $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ is close to $N(0,1)$

Consequence: For sufficiently large n:

1) $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ is approximately $N\left(\mu, \frac{\sigma^2}{n}\right)$ distributed

2) $Y = \sum_{i=1}^{n} X_i$ is approximately $N\left(n\mu, n\sigma^2\right)$ distributed

Notes: The Central Limit Theorem describes the limit behavior of the distribution of the sequence of random variables

$$W = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}, \qquad n = 1, 2, 3, \ldots$$

Note how powerful the statement is: regardless of the distribution from which we are sampling (provided it has finite mean μ and finite variance σ^2), the limiting distribution of W is $N(0, 1)$.

For sufficiently large n, the Central Limit Theorem may be used to approximate the cdf of W:

$$P(W \leq w) \approx \int_{-\infty}^{w} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = \Phi(w)$$

pdf of $N(0,1)$

In general the approximation is reasonably good for $n \geq 30$.

- In the special case of symmetric and unimodal distributions of the continuous type, $n \geq 5$ may be enough for an adequate approximation.

SUMMARY OF SAMPLING DISTRIBUTIONS

Let $X_1, X_2, \ldots, X_m$ be a random sample from $\underline{\underline{N(\mu, \sigma^2)}}$:

(1) $$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{m}\right) \Rightarrow Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{m}} \sim N(0,1)$$ ← EXACT STATEMENT

(2) $$U = \frac{(m-1)S^2}{\sigma^2} \sim \chi^2(m-1)$$

(3) $$T = \frac{Z}{\sqrt{U/(m-1)}} = \frac{\bar{X} - \mu}{S/\sqrt{m}} \sim t(m-1)$$

(4) If $Y_1, \ldots, Y_{\underline{m}}$ is an independent random sample from $N(\mu_y, \sigma_y^2)$

then $$W = \frac{\frac{(m-1)S_x^2}{\sigma_x^2} / (m-1)}{\frac{(m-1)S_y^2}{\sigma_y^2} / (m-1)} = \frac{\sigma_y^2}{\sigma_x^2} \cdot \frac{S_x^2}{S_y^2} \sim F(m-1, m-1)$$

sample size of $X_1, \ldots, X_m$

sample size of $Y_1, \ldots, Y_m$

When sampling from non - Normal distributions

CLT $\Longrightarrow$ $\dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ is approx. $N(0,1)$ for large n

Math 4501 - Probability and Statistics II

6.3 - Order statistics

↳ relevant for point estimates

Order statistics

Order statistics are observations of the random sample, arranged, or ordered, in magnitude from the smallest to the largest.

If $X_1, X_2, \dots, X_n$ are observations of a random sample of size n from a continuous-type distribution, we let the random variables

$$Y_1 < Y_2 < \cdots < Y_n$$

denote the order statistics of that sample. That is,

Y_1 is min $\rightarrow$

$$\begin{aligned} Y_1 &= \text{ smallest of } X_1, X_2, \dots, X_n \\ Y_2 &= \text{ second smallest of } X_1, X_2, \dots, X_n \\ &\vdots \\ Y_n &= \text{ largest of } X_1, X_2, \dots, X_n \end{aligned}$$

Y_n is max $\rightarrow$

Order statistics cdf

Let $Y_1 < Y_2 < \cdots < Y_n$ be the order statistics of n independent observations $X_1, X_2, \ldots, X_m$
from a distribution of the continuous type with cdf $F(x)$ and pdf $f(x) = F'(x)$.

QUESTION: cdf / pdf of Y_r ?

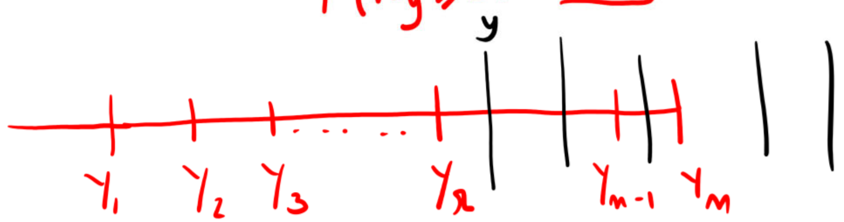
The event that the rth order statistic Y_r is at most y, $\{Y_r \leq y\}$ can occur if and only if at least r of the n observations are less than or equal to y.

The probability of "success" on each trial is then $F(y)$, and the probability of at least r successes gives

cdf of Y_r $\longrightarrow$

$$\underbrace{G_r(y)} = \underbrace{P\left(Y_r \leq y\right)}_{P(N_y \geq r)} = \sum_{k=r}^{n} \binom{n}{k} \overbrace{[F(y)]}^{p}{}^{k} \overbrace{[1 - F(y)]}^{1-p}{}^{n-k}$$

from the probability of N_y



Define N_y : r.v. representing the number of observations $X_1, X_2, \ldots, X_m$ that are less than y.

sample size

$N_y \sim \text{Binomial}(m, p)$

where $p = P(X_i \leq y) = F(y)$

Order statistics pdf

Using the cdf of the rth order statistic Y_r, note that

cdf from previous slide

$$
\begin{aligned}
\rightarrow \quad G_r(y) = P(Y_r \le y) &= \sum_{k=r}^{n} \binom{n}{k} [F(y)]^k [1 - F(y)]^{n-k} \\
&= \underbrace{\sum_{k=r}^{n-1} \binom{n}{k} [F(y)]^k [1 - F(y)]^{n-k}}_{\text{use product rule}} + [F(y)]^n .
\end{aligned}
$$

↳ term for $k = n$

Hence, the pdf of Y_r is

pdf is $g_r(y) = G'_r(y)$

$$
\begin{aligned}
g_r(y) = G_r'(y) = &\sum_{k=r}^{n-1} \binom{n}{k} (k) [F(y)]^{k-1} f(y) [1 - F(y)]^{n-k} \\
&+ \sum_{k=r}^{n-1} \binom{n}{k} [F(y)]^k (n - k) [1 - F(y)]^{n-k-1} [-f(y)] \\
&+ n [F(y)]^{n-1} f(y)
\end{aligned}
$$

$F'(y)$

Order statistics pdf

Since

$$\binom{n}{k} k = \frac{n!}{(k-1)!(n-k)!} \quad \text{and} \quad \binom{n}{k}(n-k) = \frac{n!}{k!(n-k-1)!}$$

it follows that the pdf of Y_r is

$$\boxed{g_r(y) = \frac{n!}{(r-1)!(n-r)!}[F(y)]^{r-1}[1-F(y)]^{n-r} f(y) \ , \quad a < y < b}$$

← term with k = r from the 1st sum

Note that:

1) the expression of $g_r(y)$ is the first term of the first summation in $g_r(y) = G_r'(y)$.

2) the remaining terms in $g_r(y) = G_r'(y)$ sum to zero because:

- the second term of the first summation (when $k = r + 1$) equals the negative of the first term in the second summation (when $k = r$), and so on...
- the last term of the second summation equals the negative of $n[F(y)]^{n-1} f(y)$.

Minimum and maximum order statistics pdf

The pdf of the smallest order statistic Y_1 is given by

$$g_1(y) = n[1 - F(y)]^{n-1} f(y), \quad a < y < b$$

next slide !

and the pdf of the largest order statistic Y_n is given by

$$g_n(y) = n[F(y)]^{n-1} f(y), \quad a < y < b\,.$$

below

Alternative strategy to find pdf / cdf for Y_1 and Y_n (important strategy !)

cdf of Y_n $\quad G_n(y) = P(Y_n \le y) = P(\max\{X_1, \dots, X_n\} \le y) =$

$$= P(X_1 \le y, X_2 \le y, \dots, X_n \le y) \underset{\text{independence}}{=} P(X_1 \le y) \cdot P(X_2 \le y) \dots P(X_n \le y)$$

$$= \underbrace{F(y) \cdot F(y) \dots F(y)}_{n \text{ times}} = [F(y)]^n$$

$$\Rightarrow \quad g_n(y) = G_n'(y) = n[F(y)]^{n-1} \cdot f(y)$$

cdf of Y_1 : $G_1(y) = P(Y_1 \leq y)$

$$= P(\min\{X_1, \dots, X_n\} \leq y)$$

$$= 1 - P(\min\{X_1, X_2, \dots, X_n\} > y)$$

$P(X_i \leq y) = F(y)$

$\Downarrow$

$P(X_i > y) = 1 - F(y)$

$$= 1 - P(X_1 > y, X_2 > y, \dots, X_n > y)$$

independence

$$= 1 - P(X_1 > y) \cdot P(X_2 > y) \cdots P(X_n > y)$$

$$= 1 - \underbrace{(1 - F(y)) \cdot (1 - F(y)) \cdots (1 - F(y))}_{n \text{ times}}$$

$$= 1 - [1 - F(y)]^n$$

pdf $g_1(y) = G_1'(y) = \frac{d}{dy}\left[1 - [1 - F(y)]^n\right] = -n[1 - F(y)]^{n-1} \cdot (-f(y))$

$$= n[1 - F(y)]^{n-1} \cdot f(y)$$

Example

Let $Y_1 < Y_2 < Y_3 < Y_4 < Y_5$ be the order statistics associated with $n = 5$ independent observations X_1, X_2, X_3, X_4, X_5 from the distribution with pdf

$$\Rightarrow \quad f(x) = 2x\ , \quad 0 < x < 1\ .$$

Determine the pdf of Y_4 and evaluate $P(Y_4 < 1/2)$.

Let us determine the pdf of Y_4 first:

By definition, the cdf of Y_4 is $G_4(y) = P(Y_4 \leq y)$

Note that $Y_4 \leq y$ if and only if at least 4 of the observations $X_1, X_2, \ldots, X_5$ are less or equal than y.

Let us count the number of observations $X_1, \ldots, X_5$ with values less or equal than y.

Define N_y, the random variable representing such number

$N_y \sim$ Binomial (n, p) where

$n = 5$ (# of observations $X_1, \dots, X_5$)

$p =$ prob. of success $= P(X_i \leq y) = \int_{-\infty}^{y} f(x)\,dx$

$$= \int_0^y 2x\,dx = y^2, \quad 0 \leq y \leq 1$$

Thus, we obtain that

$$G_4(y) = P(Y_4 \leq y) = P(\underbrace{N_y \geq 4}_{\substack{\text{4 or more} \\ \text{observations} \\ \text{less than } y}}) = \sum_{k=4}^{5} \binom{5}{k} \cdot (y^2)^k \cdot (1-y^2)^{5-k}$$

$$= \underset{\substack{\uparrow \\ \binom{5}{4}}}{5} (y^2)^4 (1-y^2) + 1 \cdot (y^2)^5 (1-y^2)^0$$

$$\Rightarrow G_4(y) = 5\,y^8 (1-y^2) + y^{10}$$

pdf of Y_4 is then

$$g_4(y) = G_4'(y) = \frac{d}{dy}\left[\overbrace{5y^8(1-y^2)} + \overbrace{y^{10}}\right]$$

$$= 5\cdot 8\cdot y^7(1-y^2) + \underbrace{5y^8(-2y)}_{-10y^9} + 10y^9$$

cancel out

$$= 40y^7(1-y^2)$$

← compare with what we would get using the formula

It must be the same!

To evaluate $P(Y_4 < \frac{1}{2})$:

do not do $\int_0^{1/2} g_4(y)\, dy$ (even though it works!)

it's too much effort!

Instead, use the cdf (which we also have)

$$P(Y_4 < \tfrac{1}{2}) = G_4(\tfrac{1}{2}) = 5\left(\tfrac{1}{2}\right)^8\left(1 - \left(\tfrac{1}{2}\right)^2\right) + \left(\tfrac{1}{2}\right)^{10} = \cdots$$

Math 4501 - Probability and Statistics II

6.4 - Maximum likelihood and method of moments estimation

techniques for point estimates!

Point estimation

Consider random variables for which the functional form of the pmf or pdf is known, but the distribution depends on an unknown parameter θ that may have any value in a set Ω, called the *parameter space*.

Problem: how to select precisely one member of the family

$$\{f(x, \theta), \theta \in \Omega\}$$

of pmfs / pdfs.

- determine a point estimate of the parameter θ, the value of the parameter that corresponds to the selected pdf.

Point estimator

Take a random sample from the distribution in order to gain some insight into the unknown parameter θ:

- repeat the experiment n independent times
- observe the sample, $X_1, X_2, \ldots, X_n$ ← random sample (iid) ← random variables
- try to estimate the value of θ by using the observations $x_1, x_2, \ldots, x_n$. ← real numbers

estimator in itself a r.v

The function of $X_1, X_2, \ldots, X_n$ used to estimate θ, the statistic $u(X_1, X_2, \ldots, X_n)$, is called an *estimator* of θ.

does not depend on any unknown parameter

- We would like this statistic to be such that the computed *estimate* $u(x_1, x_2, \ldots, x_n)$ is usually close to θ.

real number!

Since we are estimating one member of $\theta \in \Omega$, such an estimator is often called a *point estimator*.

Maximum likelihood estimator

→ main technique we will use to find point estimators!

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution depending on one or more unknown parameters $\theta_1, \ldots, \theta_m$.

Denote the distribution pmf or pdf by $\underbrace{f(x; \theta_1, \ldots, \theta_m)}_{\text{pdf/pmf of each of } X_1, X_2, \ldots, X_n}$, with $\overbrace{(\theta_1, \ldots, \theta_m)}^{\text{unknown parameters}} \in \Omega$.

↑ parameter space

The function

$$
\begin{aligned}
\rightarrow \quad L(\theta_1, \ldots, \theta_m) &= \prod_{i=1}^{n} f(x_i; \theta_1, \ldots, \theta_m) \\
&= \underbrace{f(x_1, \theta_1, \ldots, \theta_m)} \cdots f(x_n, \theta_1, \ldots, \theta_m), \quad (\theta_1, \ldots, \theta_m) \in \Omega,
\end{aligned}
$$

when regarded as a function of $\theta_1, \ldots, \theta_m$, is called the *likelihood function*.

$$\rightarrow L(\theta_1, \theta_2, \ldots, \theta_m) = P(X_1 = x_1, X_2 = x_2, \ldots, X_m = x_m) =$$

independency

$$= \underbrace{P(X_1 = x_1)}_{f(x_1, \theta_1, \ldots, \theta_m)} \cdot P(X_2 = x_2) \cdots \underbrace{P(X_m = x_m)}_{f(x_m, \theta_1, \ldots, \theta_m)} = \prod_{i=1}^{m} f(x_i, \theta_1, \ldots, \theta_m)$$

The functions

$$\left.\begin{aligned}\widehat{\theta_1} &= u_1\left(X_1, \ldots, X_n\right)\\ &\;\vdots\\ \widehat{\theta_m} &= u_m\left(X_1, \ldots, X_n\right)\end{aligned}\right.$$

that maximize $L(\theta_1, \ldots, \theta_m)$ are the *maximum likelihood estimators* of $\theta_1, \theta_2, \ldots, \theta_m$, respectively.

The corresponding observed values of these statistics

$$u_1\left(x_1, \ldots, x_n\right), u_2\left(x_1, \ldots, x_n\right), \ldots, u_m\left(x_1, \ldots, x_n\right)$$

are called *maximum likelihood estimates*.

Notes:
1) In many practical cases, these estimators (and estimates) are unique.
2) For many applications, there is just one unknown parameter θ. In such cases, the likelihood function is given by

$$L(\theta) = \prod_{i=1}^{n} f\left(x_i; \theta\right) = f(x_1, \theta) \cdots f(x_n, \theta)\ , \quad \theta \in \Omega\ . \leftarrow$$

like a Calculus I optimization problem!