II. Random sample

II.1 Basic definations

Def. A random sample of size N is a family of random variables

1) all having the same pdf continuous case or pmf discrete case and

2) all muturally independent.

↑

We speak about iid random variables.

↓

1) identically distributed

⚠ The condition of independence is very strong ;
We take it as a first approximation.

$\Rightarrow f_{\underline{X}}(x_1,\cdots,x_n) = f_{X_1}(x_1)\cdots f_{X_n}(x_n)$ because of independence

Def. If $Y:\mathbb{R}^N \longrightarrow \mathbb{R}^d$ is Boral measurable ($\Leftarrow$ continuous)

then $Y(X_1,\cdots,X_N)$ is a real $\overset{\text{if}}{d=1}$ or vector $\overset{\text{if}}{d>1}$ valued random variable

↳ no dependence explicitly on other variables (μ, etc)

called a statistic.

Examples

1) Sample mean: $\overline{X} = \frac{1}{N}\sum_{j=1}^{N} X_j$

2) Sample variance: $S^2 = \boxed{\frac{1}{N-1}}\sum_{j=1}^{N}(X_j-\overline{X})^2$

3) Sample standard divation: $S = \sqrt{S^2}$

Thm. Let $\underline{X} = (X_1,\cdots,X_N)$ be a random sample, with $E(X_i)=\mu$ and $Var(X_i)=\sigma^2$.

↑ arbitrary ↑

Then $E(\overline{X}) = \mu$; $Var(\overline{X}) = \frac{\sigma^2}{N}$; $E(S^2) = \sigma^2$

(Thm 5.2.6)

A useful relation: $M_{\overline{X}}(t) = \left(M_{X_i}(t/N)\right)^N$ if M_{X_i} exist

II.2 Sample from a normal distribution

Let $X_1,\cdots,X_N$ be a random sample with $X_j \sim n(\mu,\sigma^2)$ Normal distribution of mean μ and variance σ^2

↑ having or following

with the pdf of $n(\mu,\sigma^2)$ given by

$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$$

One has $E(X_j) = \mu$ and $Var(X_j) = \sigma^2$

Propositions

1) $\overline{X}$ and S^2 are independent random variables (Thm 5.3.1)

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

3) $(N-1)S^2/\sigma^2$ has a chi squared distribution with parameter $N-1$. (χ^2_{N-1})

Remark: If $Y \sim n(\mu, \sigma^2)$, then $\frac{Y-\mu}{\sigma} \sim n(0,1)$ re-scaling. Thus

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

If we compute $\sqrt{N}\,\frac{\bar{X}-\mu}{S} = \frac{(\bar{X}-\mu)/(\sigma/\sqrt{N})}{\sqrt{S^2/\sigma^2}}$ $\begin{matrix} \to \sim n(0,1) \\ \to \sim \chi^2_{N-1}/N-1 \end{matrix}$

The ratio is the student's distribution ($\equiv$ t-distribution) and does not depend on σ.

$\rightsquigarrow$ We can deduce μ.

II.3 Order statistics

Def. The order statistics of a random sample $X_1, \cdots, X_N$ are the sample values placed in ascending order:

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(N)}.$$

Examples: $X_{(1)} = \min\limits_{j \in \{1,\cdots,N\}} X_j$; $\cdots$; $X_{(N)} = \max\limits_{j \in \{1,\cdots,N\}} X_j$

Based on them one has

- Sample range: $R := X_{(N)} - X_{(1)}$
- Sample median: $M := \begin{cases} X_{(\frac{N+1}{2})} & \text{if } N \text{ odd} \\ \frac{1}{2}(X_{(N/2)} + X_{(N/2+1)}) & \text{if } N \text{ even} \end{cases}$

⚠ Sample median $\neq$ sample mean

- For $p \in (0,1)$ we define the $100p^{th}$ sample percentile to be the observation s.t. Np observations are smaller and $N(1-p)$ are larger

$$= \begin{cases} X_{(\{Np\})} & \text{if } \frac{1}{2N} < p < \frac{1}{2} \\ X_{(N+1-\{N(1-p)\})} & \text{if } \frac{1}{2} < p < 1 - \frac{1}{2N} \end{cases} \quad \text{with}$$

$$i - \tfrac{1}{2} \leq b < i + \tfrac{1}{2} \text{ with } i \in \mathbb{N} \Rightarrow \{b\} := i \quad \text{nearest integer}$$

One often uses

lower quartile := 25th sample percentile

upper quartile := 75th sample percentile