不偏分散はなぜ$n-1$で割るのか？

説明①

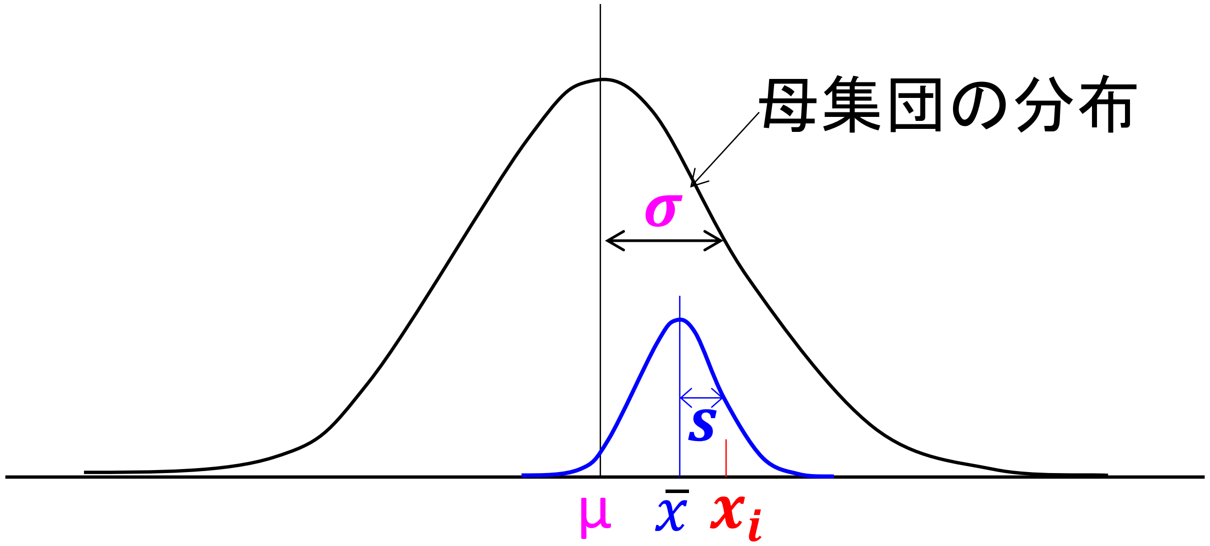
既知　　　　　　　　　　　　　不明

$$標本分散 s^2 = \frac{\sum (x_i - \bar{x})^2}{n} \leq 母分散 \sigma^2 = \frac{\sum (x_i - \mu)^2}{n}$$

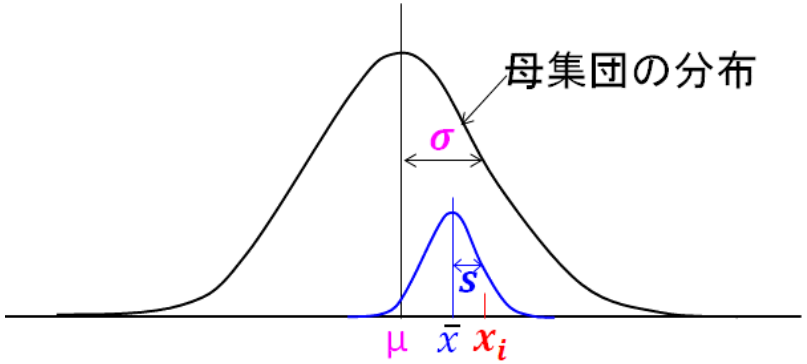
過小評価分を補正 ⇩

$$不偏分散 s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

説明②　数学の自由度より
n−1の数値がわかれば平均値の算出式から残りの1個は自明であるので自由度はn−1



標本について

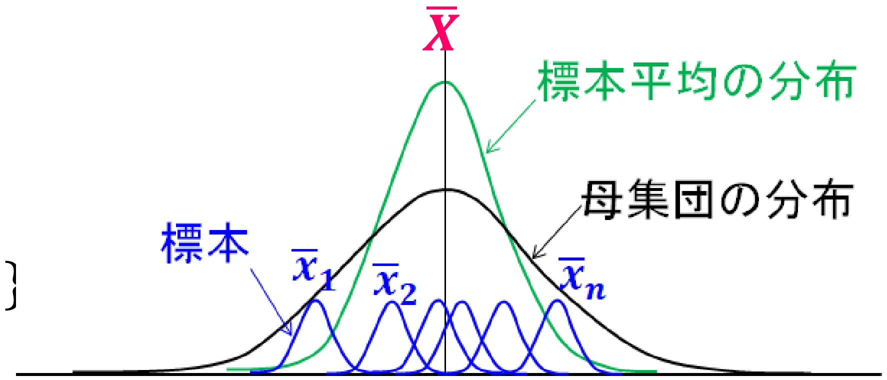


標本平均

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$$

標本分散

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$



標本平均の期待値（平均）＝母集団の平均

$$E[\bar{x}] = E\left[\frac{1}{n}(\overline{x_1} + \overline{x_2} + \cdots + \overline{x_n})\right] = \frac{1}{n}\{E[\overline{x_1}] + E[\overline{x_2}] + \cdots + E[\overline{x_n}]\}$$

$$= \frac{1}{n}(\mu + \mu + \cdots + \mu) = \mu$$

$\overline{x_i}$は母集団の構成員でもある

標本平均の分散の期待値（平均）

$$E[(\bar{X} - \mu)^2] = E\left[\left\{\frac{1}{n}(\overline{x_1} + \overline{x_2} + \cdots + \overline{x_n} - n\mu)\right\}^2\right] = E\left[\left\{\frac{(\overline{x_1} - \mu) + (\overline{x_2} - \mu) + \cdots + (\overline{x_n} - \mu)}{n}\right\}^2\right]$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} E[(\overline{x_i} - \mu)^2] = \frac{1}{n^2}(\sigma^2 + \sigma^2 + \cdots + \sigma^2) = \frac{1}{n}\sigma^2$$

標本分散

$$s^2 = \frac{1}{n}\{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\cdots+(x_n-\bar{x})^2\}$$

$$= \frac{1}{n}\{(x_1-\mu-\bar{x}+\mu)^2+(x_2-\mu-\bar{x}+\mu)^2+\cdots+(x_n-\mu-\bar{x}+\mu)^2\}$$

$$= \frac{1}{n}\{(x_1-\mu)^2-2(x_1-\mu)(\bar{x}-\mu)+(\bar{x}-\mu)^2+\cdots+(x_n-\mu)^2-2(x_n-\mu)(\bar{x}-\mu)+(\bar{x}-\mu)^2\}$$

$$= \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2 -2\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)(\bar{x}-\mu) +(\bar{x}-\mu)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2 -2\frac{1}{n}(\bar{x}-\mu)\sum_{i=1}^{n}(x_i-\mu) + (\bar{x}-\mu)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2 -2(\bar{x}-\mu)(\bar{x}-\mu) + (\bar{x}-\mu)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2 -2(\bar{x}-\mu)^2 + (\bar{x}-\mu)^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2 - (\bar{x}-\mu)^2$$

標本数nを増やせば母分散σ^2に収束

標本分散の平均=母分散−標本平均の分散

$$E[s^2] = \frac{1}{n}\sum_{i=1}^{n}E\left[(x_i-\mu)^2\right] - E[(\bar{x}-\mu)^2] = \sigma^2 - E[(\bar{x}-\mu)^2] = \boxed{\sigma^2 - \frac{1}{n}\sigma^2} = \frac{n-1}{n}\sigma^2$$

母分散

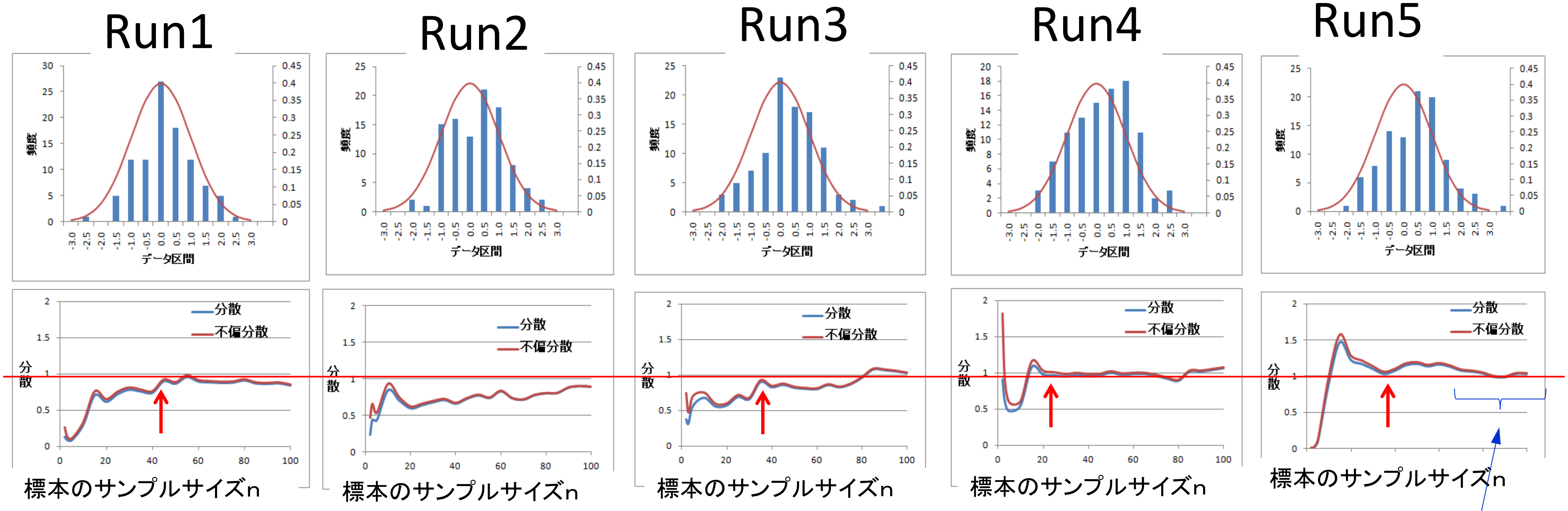
$$\sigma^2 = \frac{n}{n-1}E[S^2] = \frac{n}{n-1}\cdot\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \boxed{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

← 不偏分散で母分散を推定

不偏分散

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

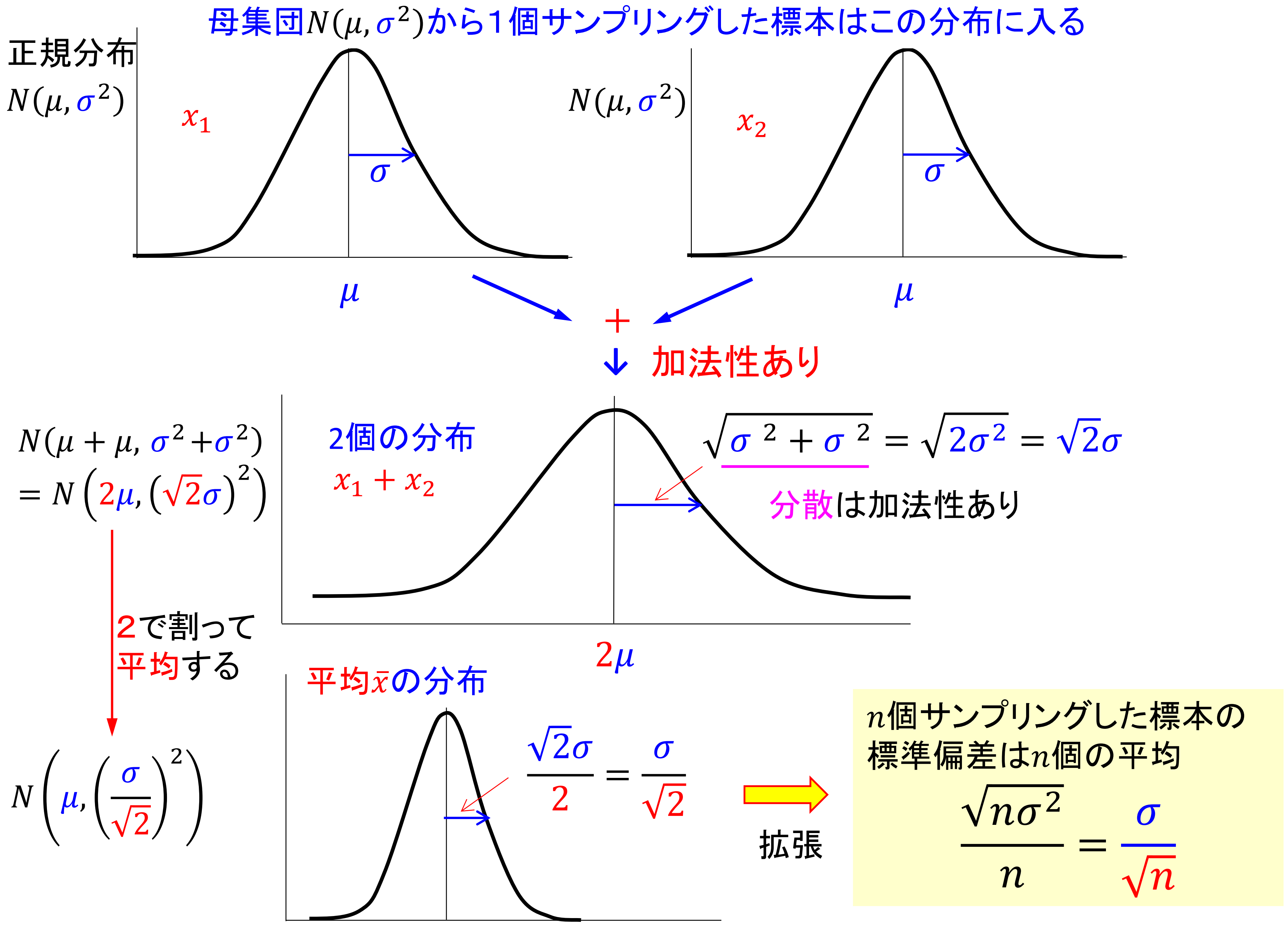
平均0、標準偏差1（分散1）の正規分布ライクの100個のデータを
−3.0～3.0の範囲で乱数で発生（Run1～5の5回実施）



標本分散$s^2 = \dfrac{\sum(x_i - \bar{x})^2}{n}$

不偏分散$s^2 = \dfrac{\sum(x_i - \bar{x})^2}{n-1}$

$\dfrac{n}{n-1}$はn増しで1に収束



母集団$N(\mu,\sigma^2)$から1個サンプリングした標本はこの分布に入る
正規分布
$N(\mu,\sigma^2)$
x_1
σ
μ
$N(\mu,\sigma^2)$
x_2
σ
μ
+
加法性あり
$N(\mu+\mu,\sigma^2+\sigma^2)$
$=N\left(2\mu,(\sqrt{2}\sigma)^2\right)$
2個の分布
x_1+x_2
$\sqrt{\sigma^2+\sigma^2}=\sqrt{2\sigma^2}=\sqrt{2}\sigma$
分散は加法性あり
2μ
2で割って
平均する
平均$\bar{x}$の分布
$N\left(\mu,\left(\frac{\sigma}{\sqrt{2}}\right)^2\right)$
$\frac{\sqrt{2}\sigma}{2}=\frac{\sigma}{\sqrt{2}}$
拡張
n個サンプリングした標本の
標準偏差はn個の平均
$\frac{\sqrt{n\sigma^2}}{n}=\frac{\sigma}{\sqrt{n}}$