◆ 標本（確率変数）の計算（１）と大数の法則

分布関数 F から抽出した標本を X と表す．この時，X は確率変数である．
同一の分布関数 F から独立に抽出した標本（サイズ＝ N ）を X_i と表す．

この時，『同一分布から抽出』という観点から，

1) $$E(X_i) = E(X_j) = E(X) = \int X\, dF = \mu$$

という関係が成立する．ただし，μ は分布関数 F の平均である．
また，1) の関係は，次式のように変形できる．

1’) $$E(X_i - \mu) = E(X_j - \mu) = E(X - \mu) = 0$$

さらに，『独立に抽出』という観点から，

確率変数 X_i **と** X_j **の同時分布** $F_{i,j}\left(X_i,\, X_j\right)$ **は，それぞれの分布** F_i **および** F_j **の積で表される．**

2)
$$F_{i,j}\left(X_i,\, X_j\right) = F_i\left(X_i\right)F_j\left(X_j\right)$$

あるいは，密度関数 $f_{i,j}$**，** f_i **および** f_j **を用いて，**

2’)
$$dF_{i,j} = f_{i,j}\left(x_i, x_j\right)dx_i dx_j = f_i\left(x_i\right)f_j\left(x_j\right)dx_i dx_j = dF_i dF_j$$

となる．したがって，以下が成立する．

3)
$$\begin{aligned} E\left(X_i - \mu\right)\left(X_{j(\neq i)} - \mu\right) &= \int\left(X_i - \mu\right)\left(X_{j(\neq i)} - \mu\right)dF_{i,j} \\ &= \int\left(X_i - \mu\right)dF_i \int\left(X_j - \mu\right)dF_j \\ &= E\left(X_i - \mu\right)E\left(X_j - \mu\right) = 0 \end{aligned}$$

（分布 F_i **と** F_j **が同一であれば，1) が成立する）**

ただし，3) において，$j = i$ である場合には，

4) $$E(X_i - \mu)(X_{j(=i)} - \mu) = \int (X_i - \mu)^2 \, dF_i = E(X_i - \mu)^2 = \sigma^2$$

となることに注意する．一方，3) と 4) の左辺は，

5) $$\begin{aligned} E(X_i - \mu)(X_j - \mu) &= E\left\{X_i X_j - \mu(X_i + X_j) + \mu^2\right\} \\ &= E(X_i X_j) - 2\mu E(X) + \mu^2 \\ &= E(X_i X_j) - \mu^2 \end{aligned}$$

である．3), 4) および 5) を整理すれば，以下のようになる．

6) $$E(X_i X_j) = \begin{cases} \mu^2 & (i \neq j) \\ \mu^2 + \sigma^2 & (i = j) \end{cases}$$

標本平均 $\overline{X}$ は，以下のとおり定義される．

7) $$\overline{X} = \frac{1}{N}\sum_{i=1\sim N} X_i$$

したがって，標本平均 $\overline{X}$ に関して，以下の関係が成立する．

8) $$E\left(\overline{X}\right) = \frac{1}{N}\sum_{i=1\sim N} E\left(X_i\right) = \frac{\mu}{N}\sum_{i=1\sim N} 1 = \mu$$

9) $$E\left(\overline{X}^2\right) = \frac{1}{N^2} E\left(\sum_{i,j(\neq i)} X_i X_j + \sum_{i,j(\neq i)} X_i^2\right) = \frac{1}{N^2}\left\{N(N-1)E\left(X_i X_j\right) + NE\left(X_i^2\right)\right\}$$

$$= \frac{1}{N^2}\left\{N(N-1)\mu^2 + N\left(\mu^2 + \sigma^2\right)\right\} = \mu^2 + \frac{\sigma^2}{N}$$

10) $$E\left(X_i \overline{X}\right) = \frac{1}{N}\sum_{i=1\sim N} E\left(X_i \overline{X}\right) = E\left(\frac{\overline{X}}{N}\sum_{i=1\sim N} X_i\right) = E\left(\overline{X}^2\right)$$

さらに，8) および 9) を用いて，以下の重要な結果を得る．

11) $$E\left(\overline{X}-\mu\right)^2 = E\left(\overline{X}^2 - 2\mu\overline{X} + \mu^2\right) = E\left(\overline{X}^2\right) - 2\mu E\left(\overline{X}\right) + \mu^2 = \frac{\sigma^2}{N}$$

これは，**標本平均の誤差分散が標本サイズに逆比例する**ことを示している．
また，その係数は母分散 σ^2 である．

以上のの結果をチェビシェフの不等式に代入して，**大数の法則**が得られる．

12) $$(1 \geq)\ \mathrm{Prob}\left\{\left|\overline{X}_N - \mu\right|^2 < \varepsilon\right\} \geq 1 - \frac{\sigma^2}{N\varepsilon} \to 1 \quad (N \to \infty)$$

この法則は，（少なくとも，平均および分散の存在する）任意の分布関数 F に対して成立する（以上の説明にて，分布関数を特定に与えていない）．

大数の法則を精密化したものが，**CLT（中心極限定理）**である．
その証明には，モーメント母関数，あるいは，特性関数を用いる．

補足：チェビシェフの不等式（テキスト p.105）

$$\int (X-\mu)^2\, dF = \int_{(X-\mu)^2 \ge (k\sigma)^2} (X-\mu)^2\, dF + \int_{(X-\mu)^2 < (k\sigma)^2} (X-\mu)^2\, dF$$

$$\ge \int_{(X-\mu)^2 \ge (k\sigma)^2} (X-\mu)^2\, dF \ge (k\sigma)^2 \int_{(X-\mu)^2 \ge (k\sigma)^2} dF = (k\sigma)^2 \left\{ 1 - \int_{(X-\mu)^2 < (k\sigma)^2} dF \right\}$$

上式の最左辺は，分散 $\sigma^2\left(= E(X-\mu)^2\right)$ **に等しい．**
また， $\mu = E(X)$ **である．したがって，上式は以下のように整理できる．**

$$(1 \ge)\ \mathrm{Prob}\left\{ \left| X - E(X) \right|^2 < k^2 E(X-\mu)^2 \right\} \ge 1 - \frac{1}{k^2}$$

上式にて， X **に** $\overline{X}_N \left(= \sum_{i=1\sim N} X_i / N\right)$ **を代入すれば，以下のようにかける．**

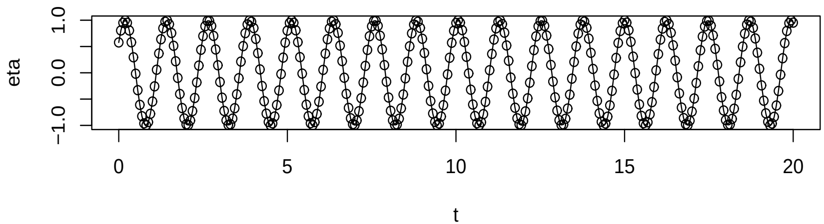
$$(1 \ge)\ \mathrm{Prob}\left\{ \left| \overline{X}_N - E(\overline{X}_N) \right|^2 < k^2 E(\overline{X}_N - \mu)^2 \right\} \ge 1 - \frac{1}{k^2}$$

また， $\overline{X}_N$ **に対する平均および分散は，** $E(\overline{X}_N) = \mu\ ;\ E(\overline{X}_N - \mu)^2 = \sigma^2/N$ **であり，**
以下のように ε **をとれば，12) が得られる．**

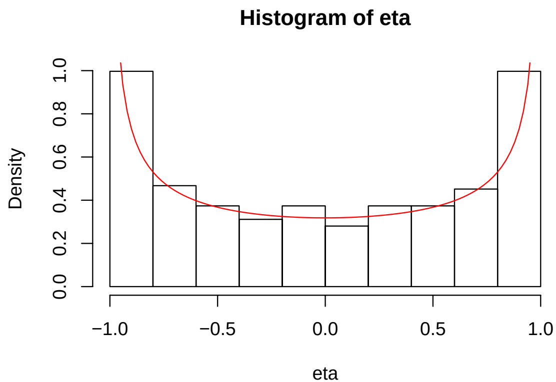
$$\varepsilon = k^2 E(\overline{X}_N - \mu)^2 = \frac{k^2\sigma^2}{N} \quad \Leftrightarrow \quad \frac{1}{k^2} = \frac{\sigma^2}{N\varepsilon}$$

◆ 中心極限定理の応用 ＝ 不規則な波（海面の水位変動）の分布

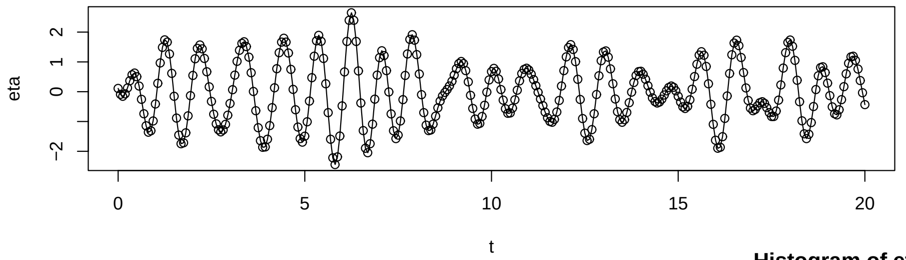
```
> t <- seq(0, 20, by=1/16)
> eta <- cos(2*pi*(t*0.810 + runif(1)))
> plot(t, eta, type="o")
```




```
> hist(eta, prob=TRUE)
> d.eta <- function(x) 1/pi/sqrt(1 - x^2)
> curve(d.eta, -1, 1, col="red", add=TRUE)
```



```
> eta <- eta*sqrt(2)
> for( freq in runif(1000 -1, 0.60, 1.40)) {
+  eta <- cos(2*pi*(t*freq + runif(1)))*sqrt(2) + eta}
> eta <- eta/sqrt(1000)
> plot(t, eta, type="o")
```




```
> hist(eta, prob=T, ylim=c(0,0.40))
> curve(dnorm, -3, 3, col="red", add=TRUE)
```