問 1

N 回コイン投げをして K 回表が出たとする（例えば 100 回投げて 55 回表が出たという状況を考える）。

このコインにいかさまが施されているかどうか判定するにはどのようにすればよいか考えよ。

解答

コインにいかさまが施されているかどうか判定するには仮説検定を行う。コインにいかさまが施されていない帰無仮説とする。このときコインを投げるごとに表・裏が出る確率は $p=0.5$ となる。また検定に用いる統計量はこの問題で N 回コイン投げをして表が出る回数 K にする。このように K は 2 項分布に従い、その確率は次式で与えられる

$$P(K=k)={}_NC_k\left(\frac{1}{2}\right)^k\left(\frac{1}{2}\right)^{N-k}={}_NC_k\left(\frac{1}{2}\right)^N$$

但し、${}_NC_k=\frac{N!}{k!(N-k)!}$ で定義される。

${}_NC_k={}_NC_{N-k}$ より $P(K=k)=P(K=N-k)$ となり分布の対称性がわかる。したがって有意水準が α であるときに

$$\sum_{k=0}^{U}{}_NC_k\left(\frac{1}{2}\right)^N<\frac{\alpha}{2}$$

を満たす最大の自然数 U を求めれば統計量 K に対して

$K<U$ または $K>N-U$ ：帰無仮説を棄却

$U\le K\le N-U$ ：帰無仮説を受容

というように仮説を検定できる。

N が十分大きいのであれば中心極限定理を用いることができ K は正規分布に従うように見なせる。また

$$E(K)=\sum_{k=0}^{N}k\,{}_NC_k\left(\frac{1}{2}\right)^N=N\sum_{k=1}^{N}{}_{N-1}C_{k-1}\left(\frac{1}{2}\right)^N$$
$$=N\sum_{k=0}^{N-1}{}_{N-1}C_k\left(\frac{1}{2}\right)^N=\frac{N}{2}\sum_{k=0}^{N-1}{}_{N-1}C_k\left(\frac{1}{2}\right)^{N-1}=\frac{N}{2}$$

$$E\left(K^2\right)=\sum_{k=0}^{N}k^2\,{}_NC_k\left(\frac{1}{2}\right)^N=\sum_{k=0}^{N}k\left[1+(k-1)\right]{}_NC_k\left(\frac{1}{2}\right)^N$$
$$=\sum_{k=0}^{N}k\,{}_NC_k\left(\frac{1}{2}\right)^N+\sum_{k=0}^{N}k(k-1)\,{}_NC_k\left(\frac{1}{2}\right)^N$$
$$=\frac{N}{2}\sum_{k=0}^{N-1}{}_{N-1}C_k\left(\frac{1}{2}\right)^{N-1}+\frac{N(N-1)}{4}\sum_{k=0}^{N-2}{}_{N-2}C_k\left(\frac{1}{2}\right)^{N-2}$$
$$=\frac{N}{2}+\frac{N(N-1)}{4}=\frac{N^2+N}{4}$$

$$V(K)=E\left(K^2\right)-\left(E(K)\right)^2=\frac{N}{4}$$

従って

$$T=\frac{(K-N/2)}{\sqrt{N/4}}=\frac{2K-N}{\sqrt{N}}$$

は標準正規分布に近づいていくので、これを用いてもよい。

以上

問 2

確率モデル

$$X=\theta+\varepsilon$$

において、n 個の独立な観測値 $X_1,X_2,\cdots,X_n$ が得られるとき、誤差 ε の分散の不偏な推定量である不偏分散が

$$\hat{\sigma}^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2$$

になることを証明せよ。ただし $\bar{X}$ が観測値の標本平均とする。

解答

まず誤差 ε の分散を σ^2 と定義する。

$$\sigma^2=V(\varepsilon)=V(X)$$

標本平均の分散は X_i の独立性より

$$V\left(\bar{X}\right)=E\left(\left(\bar{X}-\theta\right)^2\right)=E\left(\left(\frac{1}{n}\sum_{i=1}^{n}X_i-\theta\right)^2\right)$$
$$=E\left(\frac{1}{n^2}\left[\sum_{i=1}^{n}\left(X_i-\theta\right)\right]^2\right)=\frac{nV\left(\bar{X}\right)}{n^2}=\frac{\sigma^2}{n}$$

従って

$$E\left(\hat{\sigma}^2\right)=E\left(\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2\right)=\frac{1}{n-1}E\left(\sum_{i=1}^{n}\left[\left(X_i-\theta\right)-\left(\bar{X}-\theta\right)\right]^2\right)$$
$$=\frac{1}{n-1}\left[\sum_{i=1}^{n}\left[E\left(\left(X_i-\theta\right)^2\right)-E\left(\left(\bar{X}-\theta\right)^2\right)\right]\right]$$
$$=\frac{1}{n-1}\left[nV(X)-nV\left(\bar{X}\right)\right]=\frac{1}{n-1}\left[n\sigma^2-\sigma^2\right]=\sigma^2$$

$\hat{\sigma}^2$ の不偏性は明らかである。

直観的には、分散の最もよい推定量は標本分散のように思えるが、分母が n ではなく $(n-1)$ としないと不偏とならないことに注意する。

以上