Lecture 21

Covariance协方差

Definition

$$Cov(X,Y) = E((X-EX)(Y-EY))$$

$= E(XY) - E(X)E(Y)$

since $= E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y)$

Property

1. $Cov(X,X) = Var(X)$
2. $Cov(X,Y) = Cov(Y,X)$
3. $Cov(X,c) = 0$, if c is const.
4. $Cov(cX,Y) = cCov(X,Y)$
5. $Cov(X,Y+Z) = Cov(X,Y) + Cov(X,Z)$

   bilinearty(4 , 5)
6. $Cov(X+Y, Z+W) = Cov(X,Y) + Cov(X,W) + Cov(Y,Z) + Cov(Y,W)$

   $Cov(\sum_i^m a_i X_i, \sum_j^n b_j Y_j) = \sum_{i,j} a_i b_j Cov(X_i, Y_j)$
7. $Var(X_1 + X_2) = Var(X_1) + Var(X_2) + 2Cov(X_1, X_2)$

   $Var(X_1 + \ldots + X_n) = Var(X_1) + .. + Var(X_n) + 2\sum_{i<j} Cov(X_i, X_j)$

Therom

If X, Y are independent then they're uncorrelated i.e $Cov(X,Y) = 0$

Converse is false(common mistake)

Example

$$Z \sim N(0.1)$$

$X = Z, Y = Z^2, Cov(X,Y) = E(XY) - E(X)E(Y) = E(Z^3) - E(Z)E(Z^2) = 0$

but very dependent: Y is a function of X, (we know X then we know Y)

Y determines magnituide of X

Correlation相关系数

Definition

$$Corr(X,Y) = \frac{Cov(X,Y)}{SD(X)SD(Y)} = Cov(\frac{X-EX}{SD(X)}, \frac{Y-EY}{SD(Y)})$$

Therom

$-1 \leq Corr \leq 1$ (form of Cauchy-schwarz)

Proof

WLOG assume X,Y are standardized let $Corr(X,Y) = \rho$

$0 \leq Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y) = 2 + 2\rho$

$0 \leq Var(X-Y) = Var(X) + Var(Y) - 2Cov(X,Y) = 2 - 2\rho$

=> $0 \leq \rho \leq 1$

Example

Cov in a Multinomial

$(X_1, \ldots X_k) \sim Multi(n, \vec{p})$

Find $Cov(X_i, X_j)$ for all i, j

If $i = j$, $Cov(X_i, X_i) = Var(X_i) = np_i(1 - p_i)$

Now let $i \neq j$

find $Cov(X_1, X_2) = c$,

$Var(X_1 + X_2) = np_1(1 - p_1) + np_2(1 - p_2) + 2c$

$= n(p_1 + p_2)(1 - (p_1 + p_2))$ => $c = -np_1p_2$

General: $Cov(X_i, X_j) = -np_ip_j, for\ i \neq j$

Example

$X \sim Bin(n,p)$, write as $X = X_1 + \ldots X_n$, X_j are i.i.d Bern(p)

$VarX_j = EX_j^2 - (EX_j)^2 = p - p^2 = p(1-p) = pq$

Let I_A be indicator r.v. of event A

$I_A^2 = I_A, I_A^3 = I_A$

$I_A I_B = I_{A \cap B}$

$VarX = npq$ since $Cov(X_i, X_j) = 0$

Example

$X \sim HGeom(w, b, n)$

$X = X_1 + .. X_n$, $X_j = 1$ if jth ball is white; 0 otherwise

symmetry

$Var(X) = nVar(X_1) + 2\binom{n}{2}Cov(X_1, X_2)$

$Cov(X_1, X_2) = E(X_1X_2) - E(X_1)E(X_2) = \frac{w}{w+b}(\frac{w-1}{w+b-1}) - (\frac{w}{w+b})^2$