Intro Prob Lecture Notes

William Sun

April 19, 2017

Covariance

-

$$Cov(X, Y) := E(\{X - E(X)\}\{Y - E(Y)\})$$
$$= E(XY) - E(X)E(Y)$$

  - measures the "amount" of linear association between X and Y.
  - Properties:
    1) Cov(X, X) = Var(X)
    2) Cov(X, Y) = Cov(Y, X)
    3) Covariance is Bilinear

$$Cov(\sum_{i=1}^{n} a_i X_i, Y) = \sum_{i=1}^{n} a_i Cov(X_i, Y)$$

$$Cov(X, \sum_{j=1}^{m} b_j Y_j) = \sum_{j=1}^{m} b_j Cov(X, Y_j)$$

which implies

$$Cov(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j Cov(X_i, Y_j)$$

  - Combining (1) and (3), we see

-

$$
\begin{aligned}
Var(\sum_{i=1}^{n} a_i X_i) &= Cov(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{n} a_j X_i) \\
&= \sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{n} a_i a_j Cov(X_i, X_j) \\
&= \begin{bmatrix} a_1^2 Cov(X_1, X_1) & a_1 a_2 Cov(X_1, X_2) & \dots & a_1 a_n Cov(X_1, X_n) \\ a_2 a_1 Cov(X_2, X_1) & a_2^2 Cov(X_2, X_2) & \dots & a_2 a_n Cov(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ a_n a_1 Cov(X_n, X_1) & a_n a_2 Cov(X_2, X_2) & \dots & a_n^2 Cov(X_2, X_n) \end{bmatrix} \\
Var(\sum_{i=1}^{n} a_i X_i) &= \sum_{i=1}^{n} a_i^2 Var(X_i) + \sum_{i=1, j=1, i \neq j}^{n,n} a_i a_j Cov(X_i, X_j) \\
&= \sum a_i^2 Var(X_i) + 2 \sum_{i<j} a_i a_j Cov(X_i, X_j)
\end{aligned}
$$

  * The last two are called the *variance of sum formula*, important for the final!
  * Remark: In the special case where all of the random variables $X_1 \dots X_n$ are uncorrelated $(Cor(X_i, X_j) = 0 \; \forall \; i \neq j)$ then

$$
Var(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i^2 Var(X_i)
$$

  * This also holds if $X_1, X_2, \dots X - n$ are pairwise independent or independent

- Variance of $X \sim$ binomial(n, p)
  - $X = \sum_{i=1}^{n} X_i$ where $X_1, \dots X_n \sim$ independent Bernoulli(p). So
  -

$$
\begin{aligned}
Var(X) &= Var(\sum_{i=1}^{n} X_i) \\
&= \sum_{i=1}^{n} Var(X_i) \\
&= \sum_{i=1}^{n} p(1-p) \\
&= np(1-p)
\end{aligned}
$$

- $Y \sim$ hypergeometric(n, M, N)
  - $Y = \sum_{i=1}^{n} Y_i$ where $Y_i \sim$ Bernoulli($\frac{M}{N}$)
    * Y_is are not independent
  -

$$
\begin{aligned}
Var(Y) &= Var(\sum_{i=1}^{n} Y_i) \\
&= \sum_{i=1}^{n} Var(Y_i) + 2\sum_{i<j}^{n} Cov(Y_i, Y_j) \\
&\dots \\
Cov(Y_i, Y_j) &= E(Y_iY_j) - E(Y_i)E(Y_j) \\
&= 1 \cdot P(Y_i = 1, Y_j = 1) - \frac{M}{N} \cdot \frac{M}{N} \\
&= P(Y_i = 1)P(Y_j = 1|Y_i = 1) - \frac{M^2}{N^2} \\
&= \frac{M}{N} \cdot \frac{M-1}{N-1} - \frac{M^2}{N^2} \\
&= -\frac{M}{N} \cdot \frac{N-M}{N} \cdot \frac{1}{N-1} \\
&\dots \\
Var(Y) &= n \cdot \frac{M}{N} \cdot \frac{N-M}{N} + 2\Big(\frac{n(n-1)}{2}\Big)\Big(-\frac{M}{N} \cdot \frac{N-M}{N} \cdot \frac{1}{N-1}\Big) \\
&= \dots \\
&= n\frac{M}{N}\frac{N-M}{N}\frac{N-n}{N-1}
\end{aligned}
$$

    * Variance of binomial times finite population correction
      · So the variance of a hypergeometric is always less than that of its corresponding binomial
- Turns out that $-\sigma_x\sigma_y \leq Cov(X, Y) \leq \sigma_x\sigma_y$, “Cauchy-Schwarz Inequality”
  - $\rightarrow -1 \leq \rho = \frac{Cov(X,Y)}{\sigma_x\sigma_y} \leq 1$
  - ρ - correlation coefficient, measures strength of correlation, while variance measures the amount
  - If $\rho = 1$, then $Y = aX + b$ or $X = cY + d$ $a, c > 0$ if positively related

-
$$\sum_{x} x(x-1)\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

  as an exercise