Stats 426 Lecture 8

Thomas Cohn

2/15/21

Covariance and Correlation

Defn: Suppose X and Y are random variables, with means μ_X, μ_Y and standard deviations σ_X, σ_Y (respectively). Their covariance is

$$\operatorname{Cov}(X,Y) = E((X-\mu_X)(Y-\mu_Y))$$

and their correlation is

$$\operatorname{Corr}(X,Y) = \frac{\operatorname{Cov}(X,Y)}{\sigma_X\sigma_Y}$$

Some properties:

- $\operatorname{Cov}(X,Y) = E(XY) - E(X)E(Y)$
- $\operatorname{Cov}(X,Y) = \operatorname{Cov}(Y,X)$
- $\operatorname{Cov}(X,X) = \operatorname{Var}(X)$
- $X \perp Y \Rightarrow \operatorname{Cov}(X,Y) = 0$ (The converse doesn't always hold!)

Prop: Suppose $X_1, \dots, X_n, Y_1, \dots, Y_m$ are random variables, all with finite variances. Then

$$\operatorname{Cov}\left(\sum_{i=1}^n (a_i + b_iX_i), \sum_{j=1}^m (c_j + d_jY_j)\right) = \sum_{i=1}^n\sum_{j=1}^m b_id_j \operatorname{Cov}(X_i, Y_j)$$

for any constants a_i, b_i, c_j, d_j.

Some special cases:

- $\operatorname{Var}(X+Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{Cov}(X,Y)$
- $\operatorname{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \operatorname{Var}(X_i) + \sum_{i_1 \neq i_2} \operatorname{Cov}(X_{i_1}, X_{i_2})$
- $\operatorname{Var}\left(\sum_{i=1}^n (a_i + b_iX_i)\right) = \sum_{i=1}^n b_i^2 \operatorname{Var}(X_i) + \sum_{i_1 \neq i_2} b_{i_1}b_{i_2} \operatorname{Cov}(X_{i_1}, X_{i_2})$.
- $X_i \perp X_j, \forall i \neq j \Rightarrow \operatorname{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \operatorname{Var}(X_i)$

Let $\rho = \operatorname{Corr}(X,Y) = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} = \frac{\operatorname{Cov}(X,Y)}{\sigma_X\sigma_Y}$. Then

- $-1 \leq \rho \leq 1$.
  $\rho \geq -1$: $\operatorname{Var}(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}) = \operatorname{Var}(\frac{X}{\sigma_X}) + \operatorname{Var}(\frac{Y}{\sigma_Y}) + 2\operatorname{Cov}(\frac{X}{\sigma_X}, \frac{Y}{\sigma_Y}) = 1 + 1 + 2\rho = 2(1+\rho) \geq 0 \Rightarrow \rho \geq -1$
  $\rho \leq 1$: $\operatorname{Var}(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}) = \cdots = 2(1-\rho) \geq 0 \Rightarrow \rho \leq 1$.
- For bivariate normally distributed random variables X and Y, with parameters $\mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho$, we have $\operatorname{Cov}(X,Y) = \rho\sigma_X\sigma_Y$ and $\operatorname{Corr}(X,Y) = \rho$.
- $\rho = 1$ iff $\exists a \in \mathbb{R}, b > 0$ s.t. $P(Y = a + bX) = 1$.

  Proof: $\rho = 1$ iff $\operatorname{Var}(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}) = 2(1-\rho) = 0$, so let $\mu^* = E(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y})$. By the Chebyshev inequality, we have $P(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y} = \mu^*) = 1 = P(\frac{\sigma_Y}{\sigma_X}X - Y = \sigma_Y\mu^*) = P(Y = \frac{\sigma_Y}{\sigma_X}X - \sigma_Y\mu^*)$. That is, $P(Y = a + bX) = 1$ with $a = -\sigma_Y\mu^*$ and $b = \frac{\sigma_Y}{\sigma_X}$. $\square$
- Likewise, $\rho = -1$ iff $\exists a \in \mathbb{R}, b < 0$ s.t. $P(Y = a + bX) = 1$.

Conditional Expectation

Defn: (Continuous case)

$$E(Y|X=x) \stackrel{\text{def}}{=} \int\limits_{\mathbb{R}} y f_{Y|X}(y|x)\,dy$$
$$E(h(Y)|X=x) = \int\limits_{\mathbb{R}} h(y) f_{Y|X}(y|x)\,dy$$
$$\text{Var}(Y|X=x) \stackrel{\text{def}}{=} E((Y-E(Y|X=x))^2|X=x)$$

This is also defined for the discrete case, in the obvious manner.

Properties:

- $E(h(Y)|X)$ is a function of X. For a random variable X, it's also a random variable.
- $\text{Var}(Y|X=x) = E(Y^2|X=x) - E(Y|X=x)^2$
- $E(Y) = E(E(Y|X))$
- $\text{Var}(Y) = \text{Var}(E(Y|X=x)) + E(\text{Var}(Y|X=x))$

Ex: Let $X \sim \Gamma(\lambda, \alpha)$. Given $X = x$, the conditional distribution of Y is Poisson, with mean x. Then $E(Y|X) = X$ and $\text{Var}(Y|X) = X$. So

$$E(Y) = E(E(Y|X)) = E(X) = \frac{\alpha}{\lambda}$$
$$\text{Var}(Y) = \text{Var}(E(Y|X)) + E(\text{Var}(Y|X)) = \text{Var}(X) + E(X) = \frac{\alpha}{\lambda^2} + \frac{\alpha}{\lambda}$$

Note that $\text{Var}(Y) \geq \text{Var}(E(Y|X))$.

Prediction

It's common to use a function of X, $h(X)$, to predict Y.
A common criterion is "mean squared error" (MSE): $h(X) = E((Y-h(X))^2)$

$E((Y-E(Y|X))^2) \leq E((Y-h(X))^2)$, that is, the predictor $E(Y|X)$ minimizes the MSE among all predictors $h(X)$.

Moment Generating Functions

Defn: For a random variable X, its <u>moment generating funciton</u> (mgf) is $M(t) = E(e^{tX}) = \int_{\mathbb{R}} e^{tx} f(x)\,dx$.

Properties:

- For X, Y random variables with mgfs M_X, M_Y, If $M_X(t) = M_Y(t)$ for $t_1 < t < t_2$, with $0 \in (t_1, t_2)$, then X and Y have the same distribution.
- $Y = a + bX \Rightarrow M_Y(t) = e^{ta} M_X(tb)$
- $X \perp Y \Rightarrow M_{X+Y}(t) = M_X(t) M_Y(t)$, $\forall t$
- $E(X^r) = M^{(r)}(0)$ if $M^{(r)}(0)$ exists