Chapter 5 Multivariate Probability Distributions

- **Definition: Expectation under Bivariate Joint Distribution.**
  Let $g : \Omega_{XY} \to \mathbb{R}$ be a real-valued measurable function, where Ω_{XY} is the support of (X, Y). Then the expectation of $g(X, Y)$ is defined as

$$
\begin{aligned}
E[g(X,Y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)dF_{XY}(x,y) \\
&= \begin{cases} \sum\sum_{(x,y)\in\Omega_{XY}} g(x,y)f_{XY}(x,y) & \text{for d.r.v.,} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f_{XY}(x,y)\,dxdy, & \text{for c.r.v..} \end{cases}
\end{aligned}
$$

  We say that $E[g(X,Y)]$ exists if $E|g(X,Y)| < \infty$.

- **Remarks:**

  – When $g(X,Y) = X$, $E[g(X,Y)] = E(X) = \mu_X$ is the mean of X.

  – When $g(X,Y) = X^k$, $E[g(X,Y)] = E(X^k)$ is the k-th moment of X.

  – When $g(X,Y) = X^r Y^s$, $E[g(X,Y)] = E(X^r Y^s)$ is called the r-th and s-th *product moment* of X and Y.

  – When $g(X,Y) = (X - \mu_X)^2$, $E[g(X,Y)] = E(X - \mu_X)^2 = \sigma_X^2$ is the variance of X.

  – When $g(X,Y) = (X - \mu_X)(Y - \mu_Y)$,

$$\mathrm{Cov}(X,Y) \stackrel{\triangle}{=} E[g(X,Y)] = E[(X - \mu_X)(Y - \mu_Y)]$$

  is called the *covariance* of X and Y.

- **Definition: Product Moments.** The r-th and s-th order product moment of (X, Y) about the origin is define as

$$E(X^r Y^s) = \begin{cases} \sum_{x\in\Omega_X}\sum_{y\in\Omega_Y} x^r y^s f_{XY}(x,y) & \text{for d.r.v.,} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^r y^s f_{XY}(x,y)\,dxdy, & \text{for c.r.v..} \end{cases}$$

Similarly, the rth and sth central product moment is defined as

$$E\{(X-EX)^r(Y-EY)^s\}$$

$$= \begin{cases} \sum_{x\in\Omega_X}\sum_{y\in\Omega_Y}(x-\mu_X)^r(y-\mu_Y)^s f_{XY}(x,y) & \text{for d.r.v.,} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)^r(y-\mu_Y)^s f_{XY}(x,y)\,dxdy, & \text{for c.r.v..} \end{cases}$$

5.1 Expectations

- **Definition: Covariance.** Suppose $E(X^2) < \infty$ and $E(Y^2) < \infty$. Then the covariance between two variables X and Y is defined as

$$cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x - \mu_x)(y - \mu_y)dF_{XY}(x, y).$$

- **Theorem.** Suppose (X, Y) have finite second moments. Then,

$$cov(X, Y) = E(XY) - \mu_X\mu_Y.$$

- **Remarks:**
  - $\text{Cov}(X, X) = \sigma_X^2$.
  - $\text{Cov}(X, Y) = E(XY) - \mu_X\mu_Y$.
  - $E(X - \mu_X)^3 = \text{Cov}[X - \mu_X, (X - \mu_X)^2]$.
  - $E(X - \mu_X)^4 = \text{Cov}[(X - \mu_X)^2, (X - \mu_X)^2]$.
  - $\text{Cov}(aX + b, Y) = a\,\text{Cov}(X, Y)$.

- **Covariance.**
  - The covariance is a measure of the degree of co-movement between X and Y.
  - Suppose there is a high probability that large values of X tend to be observed with large values of Y, and small values of X with small values of Y, then $\text{Cov}(X, Y) > 0$.
  - On the other hand, suppose there is a high probability that large values of X tend to be observed with small values of Y, and small values of X tend to be observed with large values of Y, then $\text{Cov}(X, Y) < 0$.
  - When $\text{Cov}(X, Y) = 0$, we call X and Y are *uncorrelated*.

- **Definition:** The *correlation coefficient* between X and Y is defined as

$$\rho_{XY} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}.$$

- **Remarks:**
  - The correlation coefficient is the *standardized* covariance.
  - $|\rho_{XY}| \leq 1$, because quadratic function $\sigma_X^2 z^2 - 2\mathrm{Cov}(X,Y)z + \sigma_Y^2 \geq 0$ for all $z \in \mathbb{R}$. (**Theorem**)
  - $|\rho_{XY}| = 1$ if and only if there is a perfect linear relationship between X and Y, *i.e.*, $X = aY + b$. (**Theorem**)
  - ρ_{XY} may not capture some nonlinear relationship in X and Y.

- **Theorem:** $|\rho_{XY}| \leq 1$.

- **Theorem:** Suppose $Y = a + bX, b \neq 0$, where $\sigma_X^2 = var(X)$ exists. Then $\rho_{XY} = 1$ if $b > 0$, and $\rho_{XY} = -1$ if $b < 0$.

- **Example:** Suppose $X \sim N(0, \sigma^2)$ and $Y = X^2$. Then

$$\begin{aligned} cov(X,Y) &= E(XY) - \mu_X \mu_Y \\ &= E(X^3) = \int_{-\infty}^{\infty} x^3 \frac{1}{\sqrt{(2\pi\sigma^2)}} e^{-x^2/(2\sigma^2)} dx \\ &= 0. \end{aligned}$$

- **Linear Regression Model.**

$$Y = a + bX + \varepsilon,$$

where ε is a r.v. with $E(\varepsilon) = 0$, $var(\varepsilon) = \sigma_\varepsilon^2$, $E(X\varepsilon) = 0$. It can be shown that

$$\rho_{XY} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{b}{\sqrt{b^2 + \sigma_\varepsilon^2/\sigma_X^2}}.$$

- **Best Linear Least Squares Prediction.** Suppose X and Y are two random variables with finite second moments. Consider a linear regression function $\alpha + \beta X$ to predict Y. The Mean squared error criterion is defined as

$$\text{MSE}(\alpha, \beta) = E[Y - \alpha - \beta X]^2.$$

- **Best Linear Least Squares Prediction (continued.)** Then the optimal coefficients (α^*, β^*) that minimizes the mean squared error is

$$\beta^* = \frac{\text{cov}(X,Y)}{\text{var}(X)} = \rho_{XY}\sqrt{\frac{\text{var}(Y)}{\text{var}(X)}},$$
$$\alpha^* = \mu_Y - \beta^* \mu_X.$$

Define the prediction error as $\varepsilon = Y - (\alpha^* + \beta^* X)$. Then $E(X\varepsilon) = 0$.

- **Example: Capital Asset Pricing Model.** Let R_{pt} be the return on a portfolio during a certain time period t, r_{ft} is the risk-free interest rate, R_{mt} is the return on the market portfolio ($i.e.$, return on S&P500 index) in the same time period. The capital asset pricing model is

$$R_{pt} - r_{ft} = \beta_p (R_{mt} - r_{ft}) + \varepsilon_{pt}.$$

- **Remark.**
  - $R_{pt} - r_{ft}$ is called the excess return on the portfolio in the time period;
  - $R_{mt} - r_{ft}$ is the excess return on the market portfolio in the same period;
  - ε_{pt} represents the idiosyncratic risk peculiar to the portfolio.

- **Example: Investment Beta.** In the CAPM, usually it is assumed that $R_{mt} - r_{ft}$ and ε_{pt} are independent. The *investment beta* is

$$\beta = \frac{\mathrm{Cov}(R_p - r_f, R_m - r_f)}{\mathrm{Var}(R_m - r_f)}.$$

  - If $\beta = 1$, the portfolio is equally risky to the market portfolio.
  - If $\beta > 1$, the portfolio is more risky than the market portfolio.
  - If $\beta < 1$, the portfolio is less risky than the market portfolio.

5.1 Expectations

- **Theorem:** Suppose $Z = aX + bY + c$, then

  (1) $E(Z) = aE(X) + bE(Y) + c$.

  (2) $\text{Var}(Z) = a^2\,\text{Var}(X) + b^2\,\text{Var}(Y) + 2ab\,\text{Cov}(X, Y)$.

- **Theorem.** Suppose $Y = a_0 + \sum_{i=1}^{n} a_i X_i$, where a_i are constants. Then

  (1) $E(Y) = a_0 + \sum_{i=1}^{n} a_i E(X_i)$.

  (2) $\text{var}(Y) = \sum_{i=1}^{n} a_i^2\,\text{var}(X_i) + 2\sum_{i<j} a_i a_j\,\text{cov}(X_i, X_j)$.

- **Theorem.** Suppose two random variables (X, Y) follow a bivariate normal distribution $N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$. Then the correlation coefficient $\rho_{XY} = \rho$.

- **Theorem.** Suppose (X, Y) follows bivariate Normal distribution $N(\mu_1, \mu_2; \sigma_1^2, \sigma_2^2, \rho)$, *i.e.*,

$$f_{XY}(x, y)$$

$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$

$$exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\right\},$$

then the correlation coefficient between X and Y is

$$\rho_{XY} = \frac{\mathrm{Cov}(X, Y)}{\sigma_1\sigma_2} = \rho.$$

5.2 Joint Moment Generating Function

- **Definition: Joint Moment Generating Function.** The joint MGF of (X, Y) is defined as

$$M_{XY}(t_1, t_2) = E[\exp(t_1 X + t_2 Y)]. \ -\infty < t_1, t_2 < \infty,$$

provided the expectation exists for (t_1, t_2) in some neighborhood of $(0, 0)$.

- **Remarks:**
  - The joint MGF may not exist for some joint distributions.
  - When $M_{XY}(t_1, t_2)$ exists in a neighborhood of $(0, 0)$, it can be used to uniquely characterize the joint distribution of (X, Y).
  - $M_X(t_1) = M_{XY}(t_1, 0)$, $M_Y(t_2) = M_{XY}(0, t_2)$.

5.2 Joint Moment Generating Function

- More generally, for n-dimensional random vector $(X_1, X_2, \cdots, X_n)$, we can define the joint MGF

$$M(t_1, \cdots, t_n) = E\left[exp\left(\sum_{i=1}^{n} t_i X_i\right)\right].$$

- **Theorem:** Suppose $M_{XY}(t_1, t_2)$ exists in a neighborhood of $(0, 0)$. Then

$$E(X^k Y^l) = \frac{\partial^{k+l}}{\partial t_1^k \partial t_2^l} M_{XY}(0, 0) = M_{XY}^{(k,l)}(0, 0), \quad k, l \geq 0,$$

and

$$\mathrm{cov}(X^k, Y^l) = M_{XY}^{(k,l)}(0, 0) - M_X^{(k)}(0) M_Y^{(l)}(0).$$

In particular, $\mathrm{cov}(X, Y) = M_{XY}^{(1,1)}(0, 0) - M_X^{(1)}(0) M_Y^{(1)}(0)$.

5.3 Independence

Independence on Expectations

- **Theorem:** Suppose X and Y are independent, then

$$E[h(X)q(Y)] = E[h(X)]E[q(Y)].$$

  Equivalently,

$$cov[h(X), q(Y)] = 0.$$

- **Remarks:**
  - Uncorrelated (orthogonal) random variables may not be independent.
  - Suppose $X \sim N(0, \sigma^2)$, $Y = X^2$. Then $\mathrm{Cov}(X, Y) = 0$, but X and Y are not independent.

5.3 Implications of Independence on Expectations

- **Example:** If (X, Y) are jointly normally distributed, that is,

$$
\begin{aligned}
& f_{XY}(x, y) \\
& = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \\
& \quad exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\right\}.
\end{aligned}
$$

  Then $\mathrm{Cov}(X, Y) = 0$ if and only if X and Y are independent.

- **Remark:** The joint distribution of two normal random variables may not be normal.

Independence and Moment Generating Functions.

- **Corollary.** If X and Y are independent, and their MGF's exist in a neighborhood of 0. Let $Z = X + Y$, then

$$M_Z(t) = M_X(t)M_Y(t)$$

exists in a neighborhood of 0.

- **Remark.** This property of MGF is useful in characterizing the distribution for the sum of independent random variables.

5.3 Independence and Moment Generating Functions

- **Example:** Suppose $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$ and X, Y are independent. Then

$$X \pm Y \sim N(\mu_1 \pm \mu_2, \sigma_1^2 + \sigma_2^2).$$

- **Proof.**
  - The mgfs of X and Y are

$$\begin{aligned} M_X(t) &= e^{\mu_1 t + \frac{\sigma_1^2}{2} t^2} \\ M_Y(t) &= e^{\mu_2 t + \frac{\sigma_2^2}{2} t^2} \end{aligned}$$

  - $M_{X+Y}(t) = M_X(t) M_Y(t) = e^{(\mu_1+\mu_2)t + \frac{\sigma_1^2+\sigma_2^2}{2} t^2}$.
  - $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

5.3 Independence and Moment Generating Functions

- **Example:** Suppose $X_1, \cdots, X_n$ are independent following Poisson distributions with parameters $\lambda_1, \cdots, \lambda_n$, respectively. Find the distribution of $Y = \sum_{i=1}^n X_i$.

- **Solution:** Because $M_Y(t) = \prod_{i=1}^n M_{X_i}(t) = \prod_{i=1}^n e^{\lambda_i(e^t-1)}$, Y follows Poisson distributions with parameter $\lambda = \sum_{i=1}^n \lambda_i$.

- **Example:** Suppose $X_1, \cdots, X_n$ are independent random variables having exponential distributions with the same parameter θ. Find the distribution of $Y = \sum_{i=1}^n X_i$.

- **Solution:** $M_Y(t) = \prod_{i=1}^n M_{X_i}(t) = (1-\theta t)^{-n}$. This implies that $Y \sim$ Gamma(n, θ)

5.3 Independence

Independence and Uncorrlelatedness

- **Corollary:** If X and Y are independent, then $\text{cov}(X, Y) = 0$.

- **Theorem.** Suppose (X, Y) are jointly normally distributed. Then $cov(X, Y) = 0$ if and only if X and Y are independent.

- **Theorem.** Suppose $X \sim \text{Bernoulli}(p_1)$ and $Y \sim \text{Bernoulli}(p_2)$. Then X and Y are independent if and only if $cov(X, Y) = 0$.

5.3 Independence and Uncorrelatedness

- **Problem:** Suppose for any measurable functions $h(\cdot)$ and $q(\cdot)$, we have

$$E[h(X)q(Y)] = E[h(X)]E[q(Y)].$$

Are X and Y independent?

- **Solution:** For any $a, b \in \mathbb{R}$, let $h(x) = I(x \in (-\infty, a])$, $g(y) = I(y \in (-\infty, b])$, where $I(\cdot)$ is the indicator function. Then

$$F_{XY}(a,b) = E[h(X)g(Y)] = E[h(X)]E[g(Y)] = F_X(a)F_Y(b).$$

X and Y are independent.

5.3 Independence and Uncorrelatedness

- **Theorem:** Suppose $M_{XY}(t_1, t_2)$ exists in a neighborhood of $(0, 0)$. Then (X, Y) are independent if and only if

$$M_{XY}(t_1, t_2) = M_X(t_1)M_Y(t_2)$$

for all (t_1, t_2) in the neighborhood of $(0, 0)$.

- **Theorem:** Suppose $M_{XY}(t_1, t_2)$ exists in a neighborhood of $(0, 0)$. Then (X, Y) are independent if and only if

$$\sigma_{XY}(t_1, t_2) \stackrel{\triangle}{=} \mathrm{Cov}(e^{t_1 X}, e^{t_2 Y}) = 0$$

for all (t_1, t_2) in the neighborhood of $(0, 0)$.

- **Remark:** $\sigma_{XY}(t_1, t_2)$ can be considered as a *covariance generating function* because $\sigma_{XY}^{(k,l)}(0, 0) = \mathrm{Cov}(X^k, Y^l)$.

5.3 Independence and Uncorrelatedness

- **Theorem.** Suppose $M_{XY}(t_1, t_2)$ exists for (t_1, t_2) in a neighborhood of $(0,0)$. Then
$$cov(X,Y) = \frac{\partial^2 cov(e^{t_1 X}, e^{t_2 Y})}{\partial t_1 t_2}\Big|_{t=(0,0)}.$$
Moreover, for any positive integers r, s,
$$cov(X^r, Y^s) = \frac{\partial^{r+s} cov(e^{t_1 X}, e^{t_2 Y})}{\partial t_1^r t_2^s}\Big|_{t=(0,0)}.$$

- **Theorem.** Suppose X and Y have bounded supports. Then $cov(X^r, Y^s) = 0$ for all $r, s > 0$ if and only if X and Y are independent.

5.4 Conditional Expectations

- **Definition: Conditional Expectation.** The *conditional expectation* of $g(X,Y)$ given $X=x$ is defined as

$$E[g(X,Y) \mid X=x] = \begin{cases} \sum_y g(x,y) f_{Y|X}(y \mid x) & \text{for d.r.v.,} \\ \int_{-\infty}^{\infty} g(x,y) f_{Y|X}(y \mid x) dy & \text{for c.r.v..} \end{cases}$$

- **Remarks:**
  - The conditional expectation $E[g(X,Y) \mid X=x]$ can be considered as a function of x.
  - $E[g(X,Y) \mid X]$ is a random variable, which is a function of X.
  - $E[g(X,Y)+h(X,Y) \mid X] = E[g(X,Y) \mid X] + E[h(X,Y) \mid X]$.
  - $E[g(X,Y)q(X) \mid X] = q(X)E[g(X,Y) \mid X]$.
  - If X and Y are independent, then $E[h(Y) \mid X=x] \equiv E[h(Y)]$.

5.4 Conditional Expectations

- **Theorem: Law of Iterated Expectations.**

$$\begin{aligned} E[g(X,Y)] &= E_X\{E[g(X,Y) \mid X]\} \\ &= E_Y\{E[g(X,Y) \mid Y]\}. \end{aligned}$$

- **Remarks:**
  - The inside "E" stands for the conditional expectation.
  - The law of iterated expectations provides a two stage procedure to compute an unconditional expectation.

5.4 Conditional Expectations

- **Definition: Conditional Mean.** The *conditional mean* of Y given $X = x$ is

$$E[Y \mid X = x] = \begin{cases} \sum_y y f_{Y|X}(y \mid x) & \text{for d.r.v.,} \\ \int_{-\infty}^{\infty} y f_{Y|X}(y \mid x) dy & \text{for c.r.v..} \end{cases}$$

- **Remarks:**
  - $E[Y \mid X = x]$ is the average value of Y conditional on $X = x$.
  - $E[Y \mid X = x]$ is a function of x.

- **Example:** Let X=Gender: $X = 0$ for female and $X = 1$ for male, and let Y be the wage. Then $E[Y \mid X = 0]$ is the average wage of the female, and $E[Y \mid X = 1]$ is the average wage of the male.

5.4 Conditional Expectations

- **Theorem: Mean Squared Error (MSE) Criterion.**

$$E(Y \mid X) = \arg\min_{g(\cdot)} E[Y - g(X)]^2,$$

where the minimization is over all measurable and square-integrable functions of X.

- **Proof.** Because

$$\begin{aligned} E[Y - g(X)]^2 &= E[Y - E(Y \mid X)]^2 + E[E(Y \mid X) - g(X)]^2 \\ &\quad +2\left\{[Y - E(Y \mid X)][E(Y \mid X) - g(X)]\right\} \\ &= E[Y - E(Y \mid X)]^2 + E[E(Y \mid X) - g(X)]^2 \\ &\geq E[Y - E(Y \mid X)]^2. \end{aligned}$$

- **Remarks:**
  - The theorem shows that the best predictor of Y in terms of MSE is the conditional mean $E(Y \mid X)$.
  - In many cases, one may try to solve the constrained minimization problem

$$\min_{g \in A} E[Y - g(X)]^2,$$

where

$$A = \{g(x) : g(x) = \alpha + \beta x\}.$$

This is called *linear least square* approximation. The optimal solution is $g^*(x) = \alpha^* + \beta^* x$, where

$$\begin{aligned} \beta^* &= \frac{\text{Cov}(X, Y)}{\text{Var}(X)}, \\ \alpha^* &= \mu_Y - \beta^* \mu_X. \end{aligned}$$

- **Remarks:**
  - X and Y are independent $\Rightarrow E(Y \mid X) = E(Y) \Rightarrow \text{Cov}(X, Y) = 0$.
  - Not conversely.

- **Theorem: Regression Identify.** Suppose that $E(Y \mid X)$ exists. Then there is a random variable ε such that

$$Y = E(Y \mid X) + \varepsilon$$

and

$$E(\varepsilon \mid X) = 0.$$

Here $\varepsilon = Y - E(Y \mid X)$ is called the *regression disturbance* or *regression error*.

5.4 Conditional Expectations

- **Lemma:** Suppose $Y = a + bX + \varepsilon$. Then $E(X\varepsilon) = 0$.

- **Remark:** There is an important difference between $E(\varepsilon|X) = 0$ and $E(X\varepsilon) = 0$. Although $E(\varepsilon|X) = 0$ implies $E(X\varepsilon) = 0$, the converse is not true.

- **Example.** Suppose $\varepsilon = (X^2 - 1) + u$, where X and u are independent $N(0, 1)$ r.v.'s. Then, $E(\varepsilon|X) = X^2 - 1 + E(u|X) = X^2 - 1$, where $E(u|X) = E(u) = 0$. On the other hand, $E(X\varepsilon) = E(X^3 - X + Xu) = 0$.

- **Example.** Let the joint pdf of (X, Y) be $f_{XY}(x, y) = e^{-y}$ for $0 < x < y < \infty$. Find $\mathrm{E}(Y \mid X = x)$.

5.4 Conditional Expectations

- **Example: Efficient Market Hypothesis(EMH).** Let $Y_t = P_t - P_{t-1}$ be the asset return at time t, and I_{t-1} is the information available at time $t-1$. EMH assumes

$$E(Y_t \mid I_{t-1}) = E(Y_t).$$

- **Example: Expected Shortfall and Financial Risk Management.** The value of risk $V_t(\alpha)$ at level α is

$$P[X_t < -V_t(\alpha)|I_{t-1}] = \alpha,$$

where X_t is the return on the portfolio in time period t, and I_{t-1} is the information available at time $t-1$. The expected shortfall

$$E[X_t|X_t < -V_t(\alpha)],$$

at level α is the expected loss given a crisis has occurred.

5.4 Conditional Expectations

- **Definition: Conditional Variance.** The conditional variance of Y given X is defined as

$$\begin{aligned}\operatorname{var}(Y \mid X = x) &= E\{[Y - E(Y \mid x)]^2 \mid x\} \\ &= \int [y - E(Y \mid x)]^2 \, dF_{Y|X}(y \mid x).\end{aligned}$$

- **Remarks:**
  - $\operatorname{var}(Y \mid x) = E(Y^2 \mid x) - E^2(Y \mid x)$.
  - If $\operatorname{var}(Y \mid X) = \sigma^2$ is a constant independent of X, then $\varepsilon = Y - E(Y \mid X)$ is is called a *conditionally homoskedastic disturbance* (homoskedasticity).
  - Otherwise, it is called a *conditionally heteroskedastic disturbance* (heteroskedasticity).

5.4 Conditional Expectations

- **Example:** Let the joint pdf of (X, Y) be $f_{XY}(x, y) = e^{-y}$ for $0 < x < y < \infty$. Find $\text{var}(Y \mid X = x)$.

- **Solution:**
  - The marginal pdf of X is $f_X(x) = \int_x^\infty e^{-y} dy = e^{-x}$ for $x > 0$.
  - The conditional pdf is $f_{Y|X}(y \mid x) = e^{x-y}$ for $0 < x < y < \infty$.
  - $E(Y \mid X = x) = \int_x^\infty y e^{x-y} dy = 1 + x$ for $x > 0$.
  - $E(Y^2 \mid X = x) = \int_x^\infty y^2 e^{x-y} dy = 2 + 2x + x^2$ for $x > 0$.
  - $\text{var}(Y \mid X = x) = E(Y^2 \mid X = x) - E^2(Y \mid X = x) = 1$ for $x > 0$.

- **Example: ARCH Model.** Large volatility of an asset price today tends to be followed by another large volatility tomorrow (Volatility Clustering). Engle(1982) AutoRegressive Conditional Heteroskedasticity (ARCH) model. ARCH(1)

$$var(Y_t | I_{t-1}) = \alpha + \beta Y_{t-1}^2,$$

where $\alpha, \beta > 0$ and I_{t-1} contains information on all past returns.

- **Theorem.** $\mathrm{var}(Y \mid X) = E(Y^2 \mid X) - E^2(Y \mid X)$.

- **Example.** Suppose $Y = Z\sqrt{1+X^2}$, where Z is a r.v. with mean 0 and variance 1, and is independent of X. Find (1) $E(Y|X)$; (2) $var(Y|X)$.

- **Solution:**

$$\begin{aligned}(1)\ E(Y|X) &= E(Z\sqrt{1+X^2}|X) = \sqrt{1+X^2}E(Z|X)\\ &= \sqrt{1+X^2}E(Z) = 0 = E(Y).\end{aligned}$$

$$\begin{aligned}(2)\ var(Y|X) &= E(Y^2|X) - [E(Y|X)]^2 = E(Y^2|X)\\ &= (1+X^2)E(Z^2|X) = (1+X^2)E(Z^2) = 1+X^2.\end{aligned}$$

5.4 Conditional Expectations

- **Theorem: Variance Decomposition.** For any two random variables X and Y with finite second moments,

$$\mathrm{var}(Y) = E[\mathrm{var}(Y \mid X)] + \mathrm{var}[E(Y \mid X)].$$

- **Remark:** This theorem implies

$$\mathrm{var}(Y) \geq \mathrm{var}[E(Y \mid X)].$$

- **Higher order conditional moments:** Let $\varepsilon = Y - E(Y \mid X)$.
  - Conditional Skewness:

$$S(Y \mid X) = \frac{E(\varepsilon^3 \mid X)}{[\mathrm{var}(\varepsilon \mid X)]^{3/2}}.$$

  - Conditional Kurtosis:

$$K(Y \mid X) = \frac{E(\varepsilon^4 \mid X)}{[\mathrm{var}(\varepsilon \mid X)]^2}.$$

5.4 Conditional Expectations

- **Example:** Suppose (X, Y) follow a bivariate normal distribution

$$f_{XY}(x, y)$$
$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$
$$exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\right\}$$

- – $f_{X|Y}(x \mid y) \sim N\left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y-\mu_2), \sigma_1^2(1-\rho^2)\right)$.
  – $E(X \mid y) = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y-\mu_2)$.
  – $\mathrm{Var}(X \mid y) = \sigma_1^2(1-\rho^2)$.
  – $S(X \mid y) = 0$.
  – $K(X \mid y) = 3$.

5.4 Conditional Expectations

- **Example: Mixture Normal Distribution** Consider a case where $X = i$ with probability p_i, $i = 1, \cdots, n$. The condition probability density function $f_{Y|X}(y \mid X = i) = f_i(y) \sim N(\mu_i, \sigma_i^2)$. Then the pdf of Y is

$$f_Y(y) = \sum_i p_i f_i(y).$$

The distribution of Y is called *mixture normal* distribution. We have

$$\begin{aligned} \mathrm{Var}(Y) &= E[\mathrm{Var}(Y \mid X)] + \mathrm{Var}[E(Y \mid X)] \\ &= \sum_{i=1}^{n} p_i \sigma^2 + \sum_{i=1}^{n} p_i (\mu_i - \overline{\mu})^2, \end{aligned}$$

where $\overline{\mu} = \sum_{i=1}^{n} p_i \mu_i$.

- **THE END**