$$F_{X_1,X_2}(x_1, x_2) = P\{X_1 \le x_1, X_2 \le x_2\}$$

$$= \sum_{X_1 \le x_1} \sum_{X_2 \le x_2} P(x_1, x_2)$$

where $p(x_1, x_2) = P(X_1 = x_1, X_2 = x_2)$.

The joint(bivariate) PMF uniquely defines cdf. It also is charaterized by the two properties

$$(i)\ 0 \le p(x_1, x_2) \le 1$$
$$(ii) \sum_{all x_1} \sum_{all x_2} p(x_1, x_2) = 1$$

Theorem(Prove):

For all reals, $a < b, c < d$,

$$P\{a \le x_1 \le b, c \le x_2 \le d\} = F_{x_1,x_2}(b, d) - F_{x_1,x_2}(a, d) - F_{x_1,x_2}(b, c) + F_{x_1,x_2}(a, c)$$

A random vector (x_1, x_2) with space D is of continuous type if its CDF $F_{x_1,x_2}(x_1, x_2)$ is continuous. The function can be expressed as

$$F_{x_1,x_2}(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f_{x_1,x_2}(t_1, t_2) dt_1 dt_2$$

$\forall t_1, t_2 \in \mathbb{R}$

At points of continuity of $f_{x_1,x_2}(x_1, x_2)$ we have

$$\frac{\partial^2 F_{x_1,x_2}(x_1, x_2)}{\partial x_1 x_2} = f(x_1, x_2)$$

PDF Properties

A PDF if characterized by two properties:

$$(i)\ f_{x_1,x_2}(x_1, x_2) \ge 0$$
$$(ii) \int \int_X f(x_1, x_2) dx_1 dx_2 = 1$$

$Note : P\{(X_1, X_2) \in X\}$ is the volume under the surface

$z = f_{x_1,x_2}(x_1, x_2)$ over the set X.

Theorem

For a $Rvec(x_1, x_2)$ with $F_{x_1,x_2}(x_1, x_2)$

$$(i)F(-\infty, \infty) = F(x_1, -\infty) = F(-\infty, \infty) = 0$$
$$(ii)F(\infty, \infty) = 1$$
$$(iii)forx_1 \leq a, x_2 \leq b, F(a, b) - F(a, x_2) - F(x_1, b) + F(x_1, x_2) \geq 0$$

Conditional Expectation

$$E[g(X_2|X_1 = x_1)] = \sum_{x_2} g(x_2)p(x_2|x_1)$$

if discrete and

$$E[g(X_2|X_1 = x_1)] = \int_{-\infty}^{\infty} g(x_2)f(x_2|x_1)dx_2$$

if continuous

Measure of skewness

$$\alpha_3 = \frac{\mu_3}{\mu_2^{3/2}}$$

where skewness meaures the lack of symmetry in the pdf. The third moment is related to this.

Kurtosis

$$k = \alpha_4 = \frac{\mu_4}{\mu_2^2}$$

where kurtosis meaures the peakness of flatness of distribution. The 4th moment is related to this.

5.4 - 5.6

Independent Random Variables and Functions

CDF:

$$F(x) = P(X \le x)$$

where $b \in \mathbb{R}$

$$\implies (-\infty, b] = (-\infty, a] + (a, b]$$

$$\iff F(b) = F(a) + P\{(a, b]\}$$

$$\implies P(a < X \le b) = F(b) - F(a)$$

$I_n = \{x : a - \frac{1}{n} < x \le a + \frac{1}{n}\}$,

$$P(X = a) = P(\cap I_n)$$
$$\lim_{n\to\infty} P(I_n) = \lim_{n\to\infty} \{F(a + \frac{1}{n}) - F(a - \frac{1}{n})\}$$
$$= F(a^+) - F(a^-)$$

Definition 5.8

Let Y_1 have distribution $F_1(y_1)$ and Y_2 have distribution $F_2(y_2)$, and Y_1 and Y_2 have joint distribution function $F(y_1, y_2)$. Then Y_1 and Y_2 are said to be $independent$ if and only if

$$F(y_1, y_2) = F_1(y_1)F_2(y_2)$$

for every pair of real numbers (y_1, y_2). If Y_1 and Y_2 are not independent, they are dependent.

Discrete RVec, $\mathbb{X}(or X)$

$$P(\mathbb{X} \in B) = \sum p_X(x) \ for \ B \in \mathcal{B}$$
$$p_X(x) = p(x_1, x_2, \dots, x_k) = (X_1 = x_1, X_2 = x_2, \dots, X_k = x_k)$$

In particular, $k = 4$; Bivariate

$$1.\ P\{(X_1, X_2)\ \in A\} = \sum_{(x_1,x_2)\in A} p(x_1, x_2)$$

$$2.\ \sum_{(x_1,x_2)\in\mathbb{R}^2} p(x_1, x_2) = p\{(X_1, X_2) \in \mathbb{R}^2\} = 1$$

$$3.\ E\bigg(g(X_1, X_2)\bigg) = \sum g(x_1, x_2)p(x_1, x_2)$$

Marginals(Discrete)

$$p_{X_1}(x_1) = P(X_1 = x_1) = P(X_1 = x_2, -\infty < x_2 < \infty) = \sum_{x_2\in\mathbb{R}} p(x_1, x_2)$$

$$p_{X_2}(x_2) = \sum_{x_1\in\mathbb{R}} p(x_1, x_2)$$

Conditional(Discrete)

$$P_{X_1|X_2=x_2} = P(X_1|X_2 = x_2)$$

Continuous

CDF:

$$F_X(x) = P(X_1 \le x_1, X_2 \le x_2, \dots, X_k \le x_k)$$

PDF:

$$f_X(x) = \frac{\partial^2 F(x_1, x_2, \dots, x_k)}{\partial x_1 \partial x_2 \dots \partial x_k} = fX_1, X_2, \dots, X_k(x_1, x_2, \dots, x_k)$$

and

$$P\{X \in B\} = \int_{X_1} \int_{X_2} \dots \int_{X_k} f(x_1, x_2, \dots, x_k) dx_1, dx_2, \dots, dx_k$$

Again, $k = 2$(bivariate)

$$(i)\ P\{(X_1, X_2) \in A\} = \int \int_A f(x_1, x_2) dx_1 dx_2$$

$$(ii)\ \int_{\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2) dx_1 dx_2 = 1$$

$$(iii)\ E(X_1 X_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x_1, x_2) f(x_1, x_2) dx_1 dx_2$$

Marginals(Continuous)

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_2$$

$$f_{X_2}(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_1$$

Theorem 5.4

If Y_1 and Y_2 are discrete random variables with joint probability function $p(y_1, y_2)$ and marginal probability functions $p_1(y_1)$ and $p_2(y_2)$, respectively, then Y_1 and Y_2 are independent if and only if

$$p(y_1, y_2) = p_1(y_1) p_2(y_2)$$

for all pairs of real numbers (y_1, y_2). The corollary for this is that

$$E(Y_1 Y_2) = E(Y_1) E(Y_2)$$

Multinomial Probability Distribution

Properties:

1. The experiment consists of n identical trials.
2. The outcome of each trial falls into one of k classes of cells.

3. The probability that the outcome of a single trial falls into cell i, is p_i, $i = 1, 2, \ldots, k$ and remains the same from trial to trial. Notice that $p_1 + p_2 + p_3 + \ldots + p_k = 1$.
4. The trials are independent.
5. The random variables of interest are $Y_1, Y_2, \ldots, Y_k$, where Y_i equals the number of trials for which the outcome falls into cell i. Notice that $Y_1 + Y_2 + Y_3 + \ldots + Y_k = n$.

The joint probability distribution for $Y_1, Y_2, \ldots, Y_k$ is given

$$p(y_1, y_2, \ldots, y_k) = \frac{n!}{y_1! y_2! \ldots y_k!} p_1^{y_1} p_2^{y_2} \ldots p_k^{y_k}$$

where

$$\sum_{i=1}^{k} p_i = 1 \text{ and } \sum_{i=1}^{k} y_i = n$$

Other properties of the Multinomial Distribution are

$$\begin{aligned}&1.\ E(Y_i) = np_i\\ &2.\ V(Y_i) = np_i q_i\\ &3.\ \mathrm{Cov}(Y_s, Y_t) = -np_s p_t, \text{ if } s \neq t\end{aligned}$$

Conditional Expectations

For $Jointly$ continuous,

$$E\Bigg(g(Y_1|Y_2 = y_2)\Bigg) = \int_{-\infty}^{\infty} g(y_1) f(y_1|y_2) dy_1$$

For $Jointly$ discrete,

$$E\Bigg(g(Y_1|Y_2 = y_2)\Bigg) = \sum_{\forall\, y_1} g(y_1) p(y_1|y_2)$$

Also, the following are true

$$(i)\ E(Y_1) = E\Bigg(E(Y_1|Y_2)\Bigg)$$

$$(ii)\ V(Y_1) = E\Bigg(V(Y_1|Y_2)\Bigg) + V\Bigg(E(Y_1|Y_2)\Bigg)$$

Proof of (i):

$$\begin{aligned}E(Y_1) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y_1 f(y_1, y_2) dy_1 dy_2\\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y_1 f(y_1|y_2) f_2(y_2) dy_1 dy_2\\ &= \int_{-\infty}^{\infty}\Bigg\{\int_{-\infty}^{\infty} y_1 f(y_1|y_2)\Bigg\} f_2(y_2) dy_2\\ &= \int_{-\infty}^{\infty} E(Y_1|Y_2 = y_2) f_2(y_2) dy_2\\ &= E[E(Y_1|Y_2)]\end{aligned}$$

The Covariance of Two Random Variables

If two random variables are not independent, they are dependent or somewhat related.

Definition 5.10

If Y_1 and Y_2 are two random variables with means μ_1 and μ_2, respectively, the $covariance$ of Y_1 and Y_2 is

$$\text{Cov}(Y_1, Y_2) = E[(Y_1 - \mu_1)(Y_2 - \mu_2)]$$

Theorem

If two random variables X_1 and X_2 are independent, then

$$\text{Cov}(X_1, X_2) = 0$$

The converse may not be true.

The Correlation Coefficient

$$\rho = \frac{\text{Cov}(X_1, X_2)}{\sigma_1 \sigma_2}$$

The correlation coefficient measures the strength of linearity between X_1 and X_2. Also,

$$(i)\ 0 \leq \rho \leq 1$$
$$(ii)\ \text{Corr}(X_1, X_2) = \text{Corr}(X_2, X_1)$$

Theorem

If X_1 and X_2 are two random variables and a and b are two constants, then

$$\text{Var}(aX_1 + bX_2) = a^2\ \text{Var}(X_1) + b^2\ \text{Var}(X_2) \pm 2ab\,\text{Cov}(X_1, X_2)$$

Corollary(Prove!!!):

$$(i)\ \text{Cov}(aX_1, bX_2) = ab\,\text{Cov}(X_1, X_2)$$
$$(ii)\ \text{Cov}(X_1 + Y, Y) = \text{Cov}(X_1, Y) + \text{Cov}(Y, Y)$$
$$(iii)\ \text{Cov}(X, aX + b) = \text{Cov}(X, aX) = \text{Var}(Y) = a\,\text{Var}(X)$$

5.8 The Expected Value of and Variance of Linear Functions of Random Variables

Example:

$$x_1 - x_2 \implies a_1 x_1 + a_2 x_2$$
$$\text{Comparison } a_1 = 1\ a_2 = -1 \text{ and } \sum_{i=0}^{n} a_i = 0$$
$$U_1 = a_1 x_1 + a_2 x_2 + a_3 x_3 + \ldots + a_n x_n = \sum_{i=1}^{n} a_i Y_i$$

where all Y_is are random variables.

Theorem 5.12

Let $Y_1, Y_2, \ldots, Y_n$ and $X_1, X_2, \ldots, X_m$ be random variables with $E(Y_i) = \mu_i$ and $E(X_i) = \xi_j$. Define

$$U_1 = \sum i = 1^n a_i Y_i \text{ and } U_2 = \sum_{j=1}^{m} b_j X_j$$

for constants $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_m$. Then the following hold:

$(i)\ E(U_1) = \sum_{i=1}^{n} a_i \mu_i$

$(ii)\ V(U_1) = \sum_{i=1}^{n} a_i^2 Var(Y_i) + 2 \sum \sum_{1 \le i < j \le n} a_i a_j Cov(Y_i, Y_j) \text{ where the double sum is over all}$

$(iii)\ Cov(U_1, U_2) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j Cov(Y_i, X_j).$

Proof:

$$\begin{aligned}
Var(U_1) = E[U_1 - E(U_1)]^2 &= E\bigg(\sum_{i=1}^{n} a_i Y_i - \sum_{i=1}^{n} a_i \mu_i \bigg)^2 \\
&= E\bigg(\sum_{i=1}^{n} a_i (Y_i - \mu_i) \bigg)^2 \\
&= E\bigg(\sum_{i=1}^{n} a_i^2 (Y_i - \mu_i)^2 + \sum_{i=1}^{n} \sum_{i=1}^{n} a_i a_j (Y_i - \mu_i)(Y_j - \mu_j) \bigg) \text{ for } i \neq j \\
&= \sum_{i=1}^{n} a_i^2 E(Y_i - \mu_i)^2 + \sum_{i=1}^{n} \sum_{i=1}^{n} a_i a_j E\bigg((Y_i - \mu_i)(Y_j - \mu_j) \bigg) \text{ for } i \neq j
\end{aligned}$$

By the definitions of variance and covariance, we have

$$V(U_1) = \sum_{i=1}^{n} a_i^2 Var(Y_i) + \sum_{i=1}^{n} \sum_{i=1}^{n} a_i a_j Cov(Y_i, Y_j) \text{ for } i \neq j$$

Because $Cov(Y_i, Y_j) = Cov(Y_j, Y_i)$, we can write

$$V(U_1) = \sum_{i=1}^{n} a_i^2 V(Y_i) + 2 \sum \sum_{1 - \le i < j \le n} a_i a_j Cov(Y_i, Y_j)$$

To prove (iii),

$$\begin{aligned}
Cov(U_1, U_2) &= E\bigg([U_1 - E(U_1)][U_2 - E(U_2)]\bigg) \\
&= E\Bigg[\bigg(\sum_{i=1}^{n} a_i Y_i - \sum_{i=1}^{n} a_i \mu_i\bigg)\bigg(\sum_{j=1}^{m} b_j X_j - \sum_{j=1}^{m} b_j \xi_j\bigg)\Bigg] \\
&= E\Bigg[\bigg(\sum_{i=1}^{n} a_i (Y_i - \mu_i)\bigg)\bigg(\sum_{i=1}^{m} b_j (X_j - \xi_j)\bigg)\Bigg] \\
&= E\Bigg[\sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j (Y_i - \mu_i)(X_j - \xi_j)\Bigg] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j E[(Y_i - \mu_i)(X_j - \xi_j)] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j Cov(Y_i, X_j).
\end{aligned}$$

On observing that $Cov(Y_i, Y_i) = V(Y_i)$, we can see that (ii) is a special case of (iii).

Example 5.27

Let $Y_1, Y_2, \ldots, Y_n$ be independent random variables with $E(Y_i) = \mu$ and $V(Y_i) = \sigma^2$.(These variables may denote the outcomes of n independent trials of an experiment.) Define

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

and show that $E(\overline{Y}) = \mu$ and $V(\overline{Y}) = \sigma^2/n$.

Solution:

Because $Y_1, Y_2, \ldots, Y_n$ are all independent, $Cov(Y_i, Y_j) = 0$ for all $i \neq j$. With this being said,

$$\begin{aligned}
E(\overline{Y}) &= E\bigg(\frac{1}{n}\sum_{i=1}^{n} Y_i\bigg) \\
&= \frac{1}{n}\sum_{i=1}^{n} E(\overline{Y_i}) \\
&= \frac{1}{n}\sum_{i=1}^{n} \mu \\
&= \frac{1}{n}(n\mu) \\
&= \mu
\end{aligned}$$

and

$$
\begin{aligned}
V(\overline{Y}) &= V\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} V(Y_i) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2 \\
&= \frac{1}{n^2}(n\sigma^2) \\
&= \frac{\sigma^2}{n}
\end{aligned}
$$

Conditional Variance

Define:

$$V(Y_1|Y_2 = y_2) = E(Y_1^2|Y_2 = y_2) - \left[E(Y_1|Y_2 = y_2)\right]^2$$

Let Y_1 and Y_2 denote random variables. Then

$$V(Y_1) = E[V(Y_1|Y_2)] + V[E(Y_1|Y_2)]$$

As previously indicated, $V(Y_1|Y_2 = y_2)$ is given by

$$V(Y_1|Y_2 = y_2) = E(Y_1^2|Y_2 = y_2) - \left[E(Y_1|Y_2 = y_2)\right]^2$$

and

$$E[V(Y_1|Y_2)] = E[E(Y_1^2|Y_2)] - E\left\{[E(Y_1|Y_2)]^2\right\}$$

By definition,

$$V[E(Y_1|Y_2)] = E\left\{[E(Y_1|Y_2)]^2\right\} - \left(E[E(Y_1|Y_2)]\right)^2$$

The variance of Y_1 is

$$
\begin{aligned}
V(Y_1) &= E(Y_1^2) - [E(Y_1)]^2 \\
&= E[E(Y_1^{|}Y_2)] - \left(E[E(Y_1|Y_2)]\right)^2 \\
&= E\left(E(Y_1^2|Y_2)\right) - E\left\{[E(Y_1|Y_2)]^2\right\} + E\left([E(Y_1|Y_2)]^2\right) - \left(E[E(Y_1|Y_2)]\right)^2 \\
&= E[V(Y_1|Y_2)] + V[E(Y_1|Y_2)]
\end{aligned}
$$

Theorem:

Is $V[E(X_2|X_1)] \leq V(X_2)$? True or False? The answer is True because on the left side, the inside $E(X_2|X_1)$ means average, and that average will definitely be smaller than what is on the right side.

Simple Linear Regression

Note: Conditional Expectation is regression!!!

$$Y_i = \beta_0 + \beta_1 x_i \text{ for } i = 1, 2, \ldots, n$$

This will lead to

$$E[Y|X] = \hat{\beta}_0 + \hat{\beta}_1 x$$

Example 5.32 and 5.33 in pages 286-288