The distribution of a linear combination

If $X_1, X_2, \ldots X_n$ are n random variables, then linear combination of these r.v.s is defined by

$$Y = \sum_{i=1}^{n} a_i X_i = a_1X_1 + a_2X_2 + \cdots + a_nX_n$$

Mean of Y

$$\mu_y = E(Y) = E[a_1X_1 + a_2X_2 + \cdots + a_nX_n]$$
$$= a_1E(X_1) + a_2E(X_2) + \cdots + a_nE(X_n)$$
$$= \sum_{i=1}^{n} a_i E(X_i)$$

Variance of Y

$$\sigma_Y^2 = E(Y^2) - (E(Y))^2$$
$$= E(a_1X_1 + a_2X_2 + \cdots + a_nX_n)^2$$
$$- [E(a_1X_1 + a_2X_2 + \cdots + a_nX_n)]^2$$
$$= a_1^2 \operatorname{Var}(X_1) + a_2^2 \operatorname{Var}(X_2) + \cdots + a_n^2 \operatorname{Var}(X_n)$$
$$+ \sum_{\substack{i=1}}^{n} \sum_{\substack{j=1 \\ i \neq j}}^{n} a_i a_j \operatorname{cov}(X_i, X_j)$$

If $X_1, X_2 \cdots X_n$ are independent, then

$Cov(X_i, X_j) = 0$ for all $i \neq j$,

implying

$$Var(a_1 X_1 + a_2 X_2 + \cdots + a_n X_n)$$
$$= a_1^2 Var(X_1) + a_2^2 Var(X_2) + \cdots + a_n^2 Var(X_n)$$

In general

$$Var(a_1 X_1 + a_2 X_2 + \cdots + a_n X_n)$$
$$= \sum_{j=1}^{n} \sum_{i=1}^{n} a_i a_j \, Cov(X_i, X_j)$$

and for $i = j$, $Cov(X_i, X_i) = Var(X_i)$

Ex If X_1, X_2, X_3 are three independent random variables with means $\mu_{X_1} = 2$, $\mu_{X_2} = 3$, $\mu_{X_3} = -1$ and standard deviations $\sigma_{X_1} = 1$, $\sigma_{X_2} = 1.5$ and $\sigma_{X_3} = 0.5$ then find mean & variance of $Y = 3X_1 - 2X_2 + 5X_3$

Sol^n mean of Y =

$$E(Y) = 3E(X_1) - 2E(X_2) + 5E(X_3)$$
$$= 3 \times 2 - 2 \times 3 + 5 \times (-1) = -5$$

variance of Y = $Var(Y) = 9\,Var(X_1) + 4\,Var(X_2) + 25\,Var(X_3)$

$$= 9 \times 1^2 + 4 \times 1.5^2 + 25 \times 0.5^2 = 24.25$$

Mean random variable

If $X_1, X_2, \ldots X_n$ are n random variables, then mean random variable is

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{1}{n}\left(X_1 + X_2 + \cdots + X_n\right)$$

$$\text{mean}(\bar{X}) = E(\bar{X}) = \frac{1}{n}\left[E(X_1) + \cdots + E(X_n)\right]$$

$$\Rightarrow \mu_{\bar{X}} = \sum_{i=1}^{n} E(X_i) \Big/ n$$

Variance of $\bar{X}$

$$\sigma_{\bar{X}}^2 = \text{Var}(\bar{X}) = \sum_{j=1}^{n} \sum_{i=1}^{n} \frac{1}{n^2} \text{cov}(X_i, X_j)$$

If $X_1, X_2, \ldots X_n$ are independent, then

$$\sigma_{\bar{X}}^2 = \text{Var}(\bar{X}) = \frac{1}{n^2}\left[\text{Var}(X_1) + \text{Var}(X_2) + \cdots + \text{Var}(X_n)\right]$$

$$= \sum_{i=1}^{n} \text{Var}(X_i) \Big/ n^2$$

Note

(a) If $X_1, X_2, \ldots X_n$ have same mean μ

then $\mu_{\bar{X}} = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \frac{1}{n} \sum_{i=1}^{n} \mu$

$$= \frac{1}{n} \cdot n\mu = \mu$$

since $\boxed{\mu_{X_i} = E(X_i) = \mu}$

(b) If $X_1, X_2, \ldots X_n$ are independent and have same standard deviation i.e., $\sigma_{X_i} = \sigma$, then variance of $\bar{X}$ is

$$\sigma_{\bar{X}}^2 = \frac{1}{n^2} \sum_{i=1}^{n} Var(X_i)$$

$$= \frac{1}{n^2} \sum_{i=1}^{n} \sigma^2 = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}$$

implying standard deviation of $\bar{X}$ is

$$\boxed{\sigma_{\bar{X}} = \sigma / \sqrt{n}}$$

Central limit theorem

If $X_1, X_2 \ldots X_n$ are n independent r.v.s with same mean $\mu_{X_i} = \mu$ and same variances $\sigma_{X_i}^2 = \sigma^2$, then for sufficiently large sample n $(n \geq 30)$, then mean r.v $\bar{X}$ has approximately a normal distribution with mean $\mu_{\bar{X}} = \mu$ and variance $\sigma_{\bar{X}}^2 = \sigma^2 / n$.

Furthermore, the r.v. $T_n = n\bar{X} = X_1 + X_2 + \cdots + X_n$ has approximately a normal distribution with mean $\mu_{T_n} = n\mu$ and variance $\sigma_{T_n}^2 = n\sigma^2$ (Note, better approximation for larger n)

In this case the approximated normal curve of $\bar{X}$ is

$$f_{\bar{X}}(\bar{x}) = \frac{1}{\sigma_{\bar{X}}\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{\bar{x}-\mu_{\bar{X}}}{\sigma_{\bar{X}}}\right]^2}$$

$$= \frac{1}{\frac{\sigma}{\sqrt{n}}\cdot\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\bar{x}-\mu'}{\sigma/\sqrt{n}}\right)\right]^2$$

$$= \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\bar{x}-\mu'}{\sigma/\sqrt{n}}\right)^2\right]$$

where $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$, $x_i \in X_i$, $i=1,2,\dots,n$

the approximated normal curve of $T_n = X_1 + X_2 + \dots + X_n$ is

$$f_{T_n}(t_n) = \frac{1}{\sigma_{T_n}\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{t_n-\mu_{T_n}}{\sigma_{T_n}}\right)^2\right]$$

$$= \frac{1}{\sqrt{n}\,\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{t_n-n\mu}{\sqrt{n}\,\sigma}\right)^2\right]$$

where $t_n = x_1 + x_2 + \dots + x_n$, $x_i \in X_i$, $i=1,2,\dots,n$

For $T^* = T_n$, we have

$$f_{T^*}(t) = \frac{1}{\sigma\sqrt{2n\pi}} \exp\left[-\frac{1}{2}\left(\frac{t-n\mu}{\sqrt{n}\,\sigma}\right)^2\right],$$

$t = x_1 + x_2 + \dots + x_n$, $x_i \in X_i$, $i=1,2,\dots,n$