The Case of Normal Random Variables

When the X_i's form a random sample from a normal distribution, $\overline{X}$ and T_o are both normally distributed. Here is a more general result concerning linear combinations. The proof will be given toward the end of the section.

PROPOSITION

If $X_1, X_2, \ldots, X_n$ are independent, normally distributed rv's (with possibly different means and/or variances), then any linear combination of the X_i's also has a normal distribution. In particular, the difference $X_1 - X_2$ between two independent, normally distributed variables is itself normally distributed.

Example 6.15 (Example 6.12 continued)

The total revenue from the sale of the three grades of gasoline on a particular day was $Y = 3.5X_1 + 3.65X_2 + 3.8X_3$, and we calculated $\mu_Y = 6465$ and (assuming independence) $\sigma_Y = 493.83$. If the X_i's are normally distributed, the probability that revenue exceeds 5000 is

$$\begin{aligned} P(Y > 5000) &= P\left(Z > \tfrac{5000-6465}{493.83}\right) = P(Z > -2.967) \\ &= 1 - \Phi(-2.967) = .9985 \end{aligned}$$

The CLT can also be generalized so it applies to certain linear combinations. Roughly speaking, if n is large and no individual term is likely to contribute too much to the overall value, then Y has approximately a normal distribution.

Proofs for the Case $n = 2$

For the result concerning expected values, suppose that X_1 and X_2 are continuous with joint pdf $f(x_1, x_2)$. Then

$$\begin{aligned} E(a_1X_1 + a_2X_2) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(a_1x_1 + a_2x_2)f(x_1, x_2)\,dx_1\,dx_2 \\ &= a_1\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x_1f(x_1, x_2)\,dx_2\,dx_1 + a_2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x_2f(x_1, x_2)\,dx_1\,dx_2 \\ &= a_1\int_{-\infty}^{\infty}x_1f_{X_1}(x_1)\,dx_1 + a_2\int_{-\infty}^{\infty}x_2f_{X_2}(x_2)\,dx_2 \\ &= a_1E(X_1) + a_2E(X_2) \end{aligned}$$

Summation replaces integration in the discrete case. The argument for the variance result does not require specifying whether either variable is discrete or continuous. Recalling that $V(Y) = E[(Y - \mu_Y)^2]$,

$$\begin{aligned} V(a_1X_1 + a_2X_2) &= E\{[a_1X_1 + a_2X_2 - (a_1\mu_1 + a_2\mu_2)]^2\} \\ &= E\{a_1^2(X_1 - \mu_1)^2 + a_2^2(X_2 - \mu_2)^2 + 2a_1a_2(X_1 - \mu_1)(X_2 - \mu_2)\} \end{aligned}$$

The expression inside the braces is a linear combination of the variables $Y_1 = (X_1 - \mu_1)^2$, $Y_2 = (X_2 - \mu_2)^2$, and $Y_3 = (X_1 - \mu_1)(X_2 - \mu_2)$, so carrying the E operation through to the three terms gives $a_1^2V(X_1) + a_2^2V(X_2) + 2a_1a_2\text{Cov}(X_1, X_2)$ as required.

The previous proposition has a generalization to the case of two linear combinations:

PROPOSITION
Let U and V be linear combinations of the independent normal rv's $X_1, X_2, \ldots, X_n$. Then the joint distribution of U and V is bivariate normal. The converse is also true: if U and V have a bivariate normal distribution then they can be expressed as linear combinations of independent normal rv's.

The proof uses the methods of Section 5.4 together with a little matrix theory.

Don't have a proof.

Can you prove this last proposition?