• SUMS OF RANDOM VARIABLES
  • Variance of Sums
  • Range of Variation

• DEPENDENT RANDOM VARIABLES
  • Economic Decision Example
  • Covariance and Correlation

• For next time:
  • Devore, Sections 4.1-4.3,4.6
RANDOM SUMS

\[
\begin{array}{cccc}
& y_1 & \cdots & y_m \\
x_1 & p(x_1, y_1) & \cdots & p(x_1, y_m) & p(x_1) \\
\vdots & \vdots & \ddots & \vdots \\
x_n & p(x_n, y_1) & \cdots & p(x_n, y_m) & p(x_n) \\
& p(y_1) & \cdots & p(y_m) \\
\end{array}
\]

\[
E(X + Y) = \sum_x \sum_y (x + y) p(x, y)
\]

\[
= \sum_x x \sum_y p(x, y) + \sum_y y \sum_x p(x, y)
\]

\[
= \sum_x x p(x) + \sum_y y p(y)
\]

\[
= E(X) + E(Y) \quad ( = \mu_x + \mu_y )
\]
TEMPERATURE EXAMPLE

QUESTION:

If mean daily temperature is $60^\circ F$, then what is the mean temperature in Celsius?

ANSWER: $X =$ Temperature in Fahrenheit.

$C =$ Temperature in Celsius.

$$C = \frac{5}{9} (X - 32) = aX + b \quad \text{where} \quad a = \frac{5}{9}, \quad b = -\frac{5}{9}(32).$$

$$E(C) = E(aX + b) = \sum_x (ax + b)p(x)$$

$$= a \sum_x xp(x) + b \sum_x p(x)$$

$$= aE(X) + b$$

$\Rightarrow \quad E(C) = a(60) + b = 15.56^\circ C$
MEANS AND VARIANCES OF FUNCTIONS OF RANDOM VARIABLES

The general expression for the mean of a function of a random variable can be found on page 103 in Devore (7th ed). A full derivation of this relation for discrete random variables can be given as follows. If \( g(X) \) is any function of a random variable \( X \) with distribution \( p(x) \), and if typical values of \( g \) are denoted by \( g(x) = y \), then the mean of \( g(X) \) is given in terms of \( p(x) \) by:

\[
E[g(X)] = \sum_y y \cdot p(y) = \sum_y y \cdot p\{x : g(x) = y\} = \sum_y y \cdot \sum_{x : g(x) = y} p(x) = \sum_y \sum_{x : g(x) = y} g(x) \cdot p(x)
\]

But since the double summation \( \sum_y \sum_{x : g(x) = y} \) is precisely the summation over all values of \( x \), it then follows that

\[
E[g(X)] = \sum_x g(x) \cdot p(x)
\]

Using this result, it follows that the variance of \( g(X) \) is then given in terms of \( p(x) \) by

\[
\text{var}[g(X)] = E[g(X)^2] - (E[g(X)])^2 = \sum_x g(x)^2 \cdot p(x) - (\sum_x g(x) \cdot p(x))^2
\]
EXAMPLE.

\begin{tabular}{|c|c|c|}
\hline
$x$ & $g(x)$ & $p$ \\
\hline
$x_1$ & $g_1$ & $p_1$ \\
\hline
$x_2$ & $g_1$ & $p_2$ \\
\hline
$x_3$ & $g_2$ & $p_3$ \\
\hline
\end{tabular}

\[
E(g) = g_1 p(g_1) + g_2 p(g_2)
\]

\[
= g_1 [p(x_1) + p(x_2)] + g_2 p(x_3)
\]

\[
= g(x_1) p(x_1) + g(x_2) p(x_2) + g(x_3) p(x_3)
\]

\[
= \sum_{i=1}^{3} g(x_i) p(x_i)
\]
VARIANCE OF SUMS

\[
\text{var}(X + Y) = E\left\{\left[ (X + Y) - (\mu_x + \mu_y) \right]^2 \right\}
\]

\[
= E\left\{\left[ (X - \mu_x) + (Y - \mu_y) \right]^2 \right\}
\]

\[
= E\left\{ (X - \mu_x)^2 + 2(X - \mu_x)(Y - \mu_y) + (Y - \mu_y)^2 \right\}
\]

\[
= E\left[ (X - \mu_x)^2 \right] + 2E\left[ (X - \mu_x)(Y - \mu_y) \right] + E\left[ (Y - \mu_y)^2 \right]
\]

\[
= \text{var}(X) + 2 \text{cov}(X,Y) + \text{var}(Y)
\]

where:

\[
\text{cov}(X,Y) = E[(X - \mu_x)(Y - \mu_y)]
\]

\[
= E\left[ XY - \mu_xY - X\mu_y + \mu_x\mu_y \right]
\]

\[
= E(XY) - \mu_xE(Y) - E(X)\mu_y + \mu_x\mu_y
\]

\[
= E(XY) - \mu_x\mu_y - \mu_x\mu_y + \mu_x\mu_y
\]

\[
= E(XY) - \mu_x\mu_y
\]
QUESTION: If the standard deviation of daily temperatures is 5°F, then the range of variation is

\[ RV = 60° \pm 10° F \]

What is the range of variation in Celsius?

ANSWER: \( X = \) Temperature in Fahrenheit.

\( C = \) Temperature in Celsius.

\[ C = \frac{5}{9} (X - 32) = aX + b \quad \text{with} \quad a = .556, \ b = -17.78 \]

\[ \text{var}(C) = \text{var}(aX + b) = E\left\{(aX + b) - (a\mu + b)\right\}^2 \]

\[ = E\left[a^2 (X - \mu)^2\right] = a^2 E[(X - \mu)^2] \]

\[ = a^2 \text{var}(X) \Rightarrow \sigma(C) = |a|\sigma(X) \]

\[ \Rightarrow \quad \sigma(C) = (.556)5 = 2.78° C \]

\[ \Rightarrow \quad RV = 15.56 \pm 5.56° C \]
SUMS OF RANDOM VARIABLES

• MEAN OF SUMS

\[
E(b + \sum_{i=1}^{n} a_i X_i) = b + \sum_{i=1}^{n} a_i E(X_i)
\]

• GENERAL VARIANCE OF SUMS

\[
\text{var}(b + \sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \text{cov}(X_i, X_j)
\]

\[
= \sum_{i=1}^{n} a_i^2 \text{var}(X_i) + \sum_{i=1}^{n} \sum_{j \neq i} a_i a_j \text{cov}(X_i, X_j)
\]

• INDEPENDENCE CASE

\[
\text{var}(b + \sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i^2 \text{var}(X_i)
\]
VARIANCE OF WEIGHTED SUMS

The general expressions for the mean and variance of weighted sums can be found on p.219 in Devore (7th ed). The linearity of means is straightforward, but the expression for variance is more complex. Hence we start with a full derivation for the case of two random variables. For any random variables, $X_1$ and $X_2$, with $E(X_1) = \mu_1$ and $E(X_2) = \mu_2$, and any scalars $a_1$ and $a_2$,

(1) \[ \text{var}(a_1X_1 + a_2X_2) = E\left\{ \left( (a_1X_1 + a_2X_2) - (a_1\mu_1 + a_2\mu_2) \right)^2 \right\} \]

\[ = E\left\{ \left( (a_1X_1 - a_1\mu_1) + (a_2X_2 - a_2\mu_2) \right)^2 \right\} \]

\[ = E\left\{ (a_1X_1 - a_1\mu_1)^2 + 2(a_1X_1 - a_1\mu_1)(a_2X_2 - a_2\mu_2) + (a_2X_2 - a_2\mu_2)^2 \right\} \]

\[ = E\left[ a_1^2(X_1 - \mu_1)^2 \right] + 2a_1a_2E\left[ (X_1 - \mu_1)(X_2 - \mu_2) \right] + E\left[ a_2^2(X_2 - \mu_2)^2 \right] \]

\[ = a_1^2\text{var}(X_1) + 2a_1a_2\text{cov}(X_1, X_2) + a_2^2\text{var}(X_2) \]

where,

(2) \[ \text{cov}(X_1, X_2) = E\left[ (X_1 - \mu_1)(X_2 - \mu_2) \right] \]

Since \( \text{var}(X) = E\left[ (X - \mu)(X - \mu) \right] = \text{cov}(X, X) \), expression (1) can also be written as

(3) \[ \text{var}(a_1X_1 + a_2X_2) = a_1^2\text{var}(X_1) + 2a_1a_2\text{cov}(X_1, X_2) + a_2^2\text{var}(X_2) \]

\[ = \sum_{i=1}^{2} \sum_{j=1}^{2} a_i a_j \text{cov}(X_i, X_j) \]

Finally, if $X_1$ and $X_2$ are independent random variables, then \( \text{cov}(X_1, X_2) = 0 \), so that expression (1) reduces to

(4) \[ \text{var}(a_1X_1 + a_2X_2) = a_1^2\text{var}(X_1) + a_2^2\text{var}(X_2) \]

To extend these results to the general case, observe first that the squared sum of any set of variables $(x_1, \ldots, x_n)$ is given by
Moreover, one can decompose this into squared terms plus cross product terms:

\[(\sum_{i=1}^{n} x_i)^2 = \sum_{i=1}^{n} (x_i^2 + \sum_{j \neq i} x_i x_j) = \sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n} \sum_{j \neq i} x_i x_j\]  

Using (5) we can extend (3) as follows:

\[
\text{var}\left(\sum_{i=1}^{n} a_i X_i\right) = \mathbb{E}\left[\left(\sum_{i=1}^{n} a_i X_i - \sum_{i=1}^{n} a_i \mu_i\right)^2\right] = \mathbb{E}\left[\left(\sum_{i=1}^{n} a_i (X_i - \mu_i)\right)^2\right]
\]

\[
= \mathbb{E}\left[\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j (X_i - \mu_i)(X_j - \mu_j)\right]
\]

\[
= \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)]
\]

\[
= \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \text{cov}(X_i, X_j)
\]

Then using (6), we can extend (1) as:

\[
\text{var}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \text{var}(X_i, X_j) + \sum_{i=1}^{n} \sum_{j \neq i} a_i a_j \text{cov}(X_i, X_j)
\]

\[
= \sum_{i=1}^{n} a_i^2 \text{var}(X_i) + \sum_{i=1}^{n} \sum_{j \neq i} a_i a_j \text{cov}(X_i, X_j)
\]

Finally, for the case of independent variables, (8) reduces to:

\[
\text{var}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \text{var}(X_i)
\]
DISCOUNT PROBLEM:

Best Buy is offering a discount on the joint purchase of (TV+VCR). They made 300 discount sales last year, but want to increase sales by more advertising. For an acceptable rate of return on their investment, they require an expected revenue increase of $500,000 from new sales.

Q1. How many new sales are needed?

Q2. What is the range of variation in revenue per customer?

Q3. Do higher (lower) priced TV’s and VCR’s tend to be sold together?
DISCOUNT EXAMPLE

• RANDOM VARIABLES

\[ X_1 = TV \text{ price} \]
\[ X_2 = VCR \text{ price} \]

• JOINT PROBABILITIES

\[
\begin{array}{c|c|c|c}
X_1 & 300 & 500 & 900 \\
\hline
150 & .40 & .10 & .01 \\
250 & .04 & .15 & .30 \\
\hline
.44 & .25 & .31 \\
\end{array}
\]

\[
\begin{array}{c|c|c|c|c}
X_2 & .51 & .44 \\
\hline
150 & .01 & .44 \\
250 & .04 & .44 \\
\hline
.31 & .44 \\
\end{array}
\]
QUESTION 1:

\[ R = X_1 + X_2 \]

\[ R_T = \sum_{i=1}^{n} R_i \]

\[ \Rightarrow E(R_T) = \sum_{i=1}^{n} E(R_i) = nE(R) \]

So:

\[ 500,000 \leq E(R_T) = nE(R) \Rightarrow n \geq \frac{500,000}{E(R)} \]

Solution:

\[ E(R) = E(X_1) + E(X_2) \]

\[ E(X_1) = 300(.44) + 500(.25) + 900(.31) = 536 \]

\[ E(X_2) = 150(.51) + 250(.49) = 199 \]

\[ \Rightarrow E(R) = 536 + 199 = 735 \]

\[ \Rightarrow n \geq \frac{500,000}{735} \approx 680 \text{ sales} \]
Plot Joint Distribution

\[ M_1 = 536 \]
\[ M_2 = 199 \]

Summary:

\[ \text{Figure 5.4 (Devore, p. 217)} \]

\[ \text{COV} > 0 \quad \text{COV} < 0 \quad \text{COV} = 0 \]