LECTURE 12: Sums of independent random variables;
Covariance and correlation

• The PMF/PDF of $X + Y$ ($X$ and $Y$ independent)
  - the discrete case
  - the continuous case
  - the mechanics
  - the sum of independent normals

• Covariance and correlation
  - definitions
  - mathematical properties
  - interpretation
The distribution of $X + Y$: the discrete case

- $Z = X + Y$; $X, Y$ independent, discrete

known PMFs

$p_Z(3) =$

\[ p_Z(z) = \sum_x p_X(x) p_Y(z - x) \]
Discrete convolution mechanics

\[ p_Z(z) = \sum_x p_X(x) p_Y(z-x) \]

- To find \( p_Z(3) \):
  
  - Flip (horizontally) the PMF of \( Y \)
  
  - Put it underneath the PMF of \( X \)
  
  - Right-shift the flipped PMF by \( 3 \)
  
  - Cross-multiply and add
  
  - Repeat for other values of \( z \)
The distribution of $X + Y$: the continuous case

- $Z = X + Y$; $X, Y$ independent, continuous known PDFs

Conditional on $X = x$:

Joint PDF of $Z$ and $X$:

From joint to the marginal: $f_Z(z) = \int_{-\infty}^{\infty} f_{X,Z}(x, z) \, dx$

- Same mechanics as in discrete case (flip, shift, etc.)
The sum of independent normal r.v.'s

- $X \sim N(\mu_x, \sigma_x^2)$, $Y \sim N(\mu_y, \sigma_y^2)$, independent

$Z = X + Y$

\[
f_X(x) = \frac{1}{\sqrt{2\pi \sigma_x^2}} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}}
\]

\[
f_Y(y) = \frac{1}{\sqrt{2\pi \sigma_y^2}} e^{-\frac{(y-\mu_y)^2}{2\sigma_y^2}}
\]

\[
f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) \, dx
\]

(algebra)

\[
f_Z(z) = \frac{1}{\sqrt{2\pi (\sigma_x^2 + \sigma_y^2)}} e^{-\frac{(z-\mu_x-\mu_y)^2}{2(\sigma_x^2 + \sigma_y^2)}}
\]

The sum of finitely many independent normals is normal
Covariance

- Zero-mean, discrete $X$ and $Y$
  - if independent: $E[XY] =$

Definition for general case:

$$\text{cov}(X, Y) = E \left[ (X - E[X]) \cdot (Y - E[Y]) \right]$$

- independent $\Rightarrow$ $\text{cov}(X, Y) = 0$
  (converse is not true)
Covariance properties

\[ \text{cov}(X, X) = \]

\[ \text{cov}(aX + b, Y) = \]

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

\[ \text{cov}(X, Y) = E[XY] - E[X]E[Y] \]
The variance of a sum of random variables

\[ \text{var}(X_1 + X_2) = \]
The variance of a sum of random variables

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

\[
\text{var}(X_1 + \cdots + X_n) = \\
\sum_{i=1}^{n} \text{var}(X_i) + \sum_{\{(i,j) : i \neq j\}} \text{cov}(X_i, X_j)
\]
The Correlation coefficient

- Dimensionless version of covariance:
  \[ -1 \leq \rho \leq 1 \]

- Measure of the degree of "association" between \( X \) and \( Y \)

- Independent \( \Rightarrow \rho = 0 \), "uncorrelated" (converse is not true)

- \( |\rho| = 1 \Leftrightarrow (X - \mathbb{E}[X]) = c(Y - \mathbb{E}[Y]) \) (linearly related)

- \( \text{cov}(aX + b, Y) = a \cdot \text{cov}(X, Y) \Rightarrow \rho(aX + b, Y) = \)
Proof of key properties of the correlation coefficient

\[ \rho(X, Y) = \mathbb{E}\left[ \frac{(X - \mathbb{E}[X])}{\sigma_X} \cdot \frac{(Y - \mathbb{E}[Y])}{\sigma_Y} \right] \]

\[ -1 \leq \rho \leq 1 \]

- Assume, for simplicity, zero means and unit variances, so that \( \rho(X, Y) = \mathbb{E}[XY] \)

\[ \mathbb{E}\left[ (X - \rho Y)^2 \right] = \]

If \( |\rho| = 1 \), then
Interpreting the correlation coefficient

- Association does not imply causation or influence
  
  \[ X: \text{math aptitude} \]
  
  \[ Y: \text{musical ability} \]

- Correlation often reflects underlying, common, hidden factor
  
  Assume, \( Z, V, W \) are independent
  
  \[ X = Z + V \]
  
  \[ Y = Z + W \]

  Assume, for simplicity, that \( Z, V, W \) have zero means, unit variances

\[ \rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} \]
Correlations matter...

- A real-estate investment company invests $10M in each of 10 states. At each state $i$, the return on its investment is a random variable $X_i$, with mean 1 and standard deviation 1.3 (in millions).

$$\text{var}(X_1 + \cdots + X_{10}) = \sum_{i=1}^{10} \text{var}(X_i) + \sum_{\{(i,j): i \neq j\}} \text{cov}(X_i, X_j)$$

- If the $X_i$ are uncorrelated, then:

$$\text{var}(X_1 + \cdots + X_{10}) =$$

$$\sigma(X_1 + \cdots + X_{10}) =$$

- If for $i \neq j$, $\rho(X_i, X_j) = 0.9$:

$$\sigma(X_1 + \cdots + X_{10}) =$$