Recap

- Location-scale families
- Exponential family notes
- Multiple random variables
**Example**

- Six points in the plane, all equally likely to be chosen:

  - Points \((x, y)\) with \(0 \leq x, 0 \leq y, x + y \leq 2\), and \(x, y\) integers.
  - The joint PMF is

    \[
    f(x, y) = \begin{cases} 
    \frac{1}{6} & \text{if } x, y \text{ are integers, } x, y \geq 0, x + y \leq 2 \\
    0 & \text{otherwise}
    \end{cases}
    \]
Example (continued)

• We can find

\[ P(X + Y = 2) = f(2, 0) + f(1, 1) + f(0, 2) = \frac{1}{2} \]

• \( X \) and \( Y \) are random variables, so they have PMF’s:

<table>
<thead>
<tr>
<th>( x )</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f_X(x) )</td>
<td>( \frac{1}{2} )</td>
<td>( \frac{1}{3} )</td>
<td>( \frac{1}{6} )</td>
</tr>
</tbody>
</table>

• \( f_Y(y) \) is the same by symmetry.
Example (continued)

- Sometimes we use a table to represent joint PMF’s:

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1/6</td>
<td>1/6</td>
<td>1/6</td>
</tr>
<tr>
<td>1</td>
<td>1/6</td>
<td>1/6</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1/6</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

- The PMFs of $X$ and $Y$ appear in the margins.
- They are called *marginal PMFs*. 
• In general, we can compute the marginal PMF of, say, $X$ as

$$f_X(x) = \sum_y f(x, y)$$

• For $n$ variables,

$$f_{X_1}(x_1) = \sum_{x_2, \ldots, x_n} f(x_1, \ldots, x_n)$$

• For $n \geq 3$ we can also compute a joint marginal PMF of two or more variables:

$$f_{X,Z}(x, z) = \sum_y f(x, y, z)$$
• We can compute the expected value of a random variable from its marginal distribution.

• For our example,

\[
E[X] = 0 \times \frac{1}{2} + 1 \times \frac{1}{3} + 2 \times \frac{1}{6} = \frac{2}{3}
\]

• In general, the expectation of a function \( g \) of several discrete random variables can be computed as

\[
E[g(X_1, \ldots, X_n)] = \sum_{x_1, \ldots, x_n} g(x_1, \ldots, x_n) f(x_1, \ldots, x_n)
\]

• You get the same answer using this approach:

\[
E[X] = 0 \times \frac{1}{6} + 0 \times \frac{1}{6} + 0 \times \frac{1}{6} + 1 \times \frac{1}{6} + 1 \times \frac{1}{6} + 2 \times \frac{1}{6} = \frac{2}{3}
\]
This is always true:

\[
E[g(X)] = \sum_x \sum_y g(x) f(x, y)
\]

\[
= \sum_x g(x) \left( \sum_y f(x, y) \right)
\]

\[
= \sum_x g(x) f_X(x)
\]

We can also now verify that \( E[X + Y] = E[X] + E[Y] \):

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

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

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

\[
= E[X] + E[Y]
\]
Jointly Continuous Random Variables

Random variables $X_1, \ldots, X_n$ are jointly continuous if there is a non-negative function $f$ such that

$$P((X_1, \ldots, X_n) \in A) = \int_A \cdots \int f(x_1, \ldots, x_n) \, dx_1 \cdots dx_n$$

for “all” $A \subset \mathbb{R}^n$.

- $f$ is a joint PDF of $X_1, \ldots, X_n$
- Any non-negative function $f(x_1, \ldots, x_n)$ with
  $$\int \cdots \int f(x_1, \ldots, x_n) \, dx_n \cdots dx_1 = 1$$
  is a joint PDF.
- It is possible for $X, Y$ to be marginally continuous but not jointly continuous.
Example

- Let \( A = \{(x, y) : x \geq 0, y \geq 0, x + y \leq 1\} \)

- The area of \( A \) is \(|A| = \frac{1}{2}\).

- Let

\[
  f(x, y) = \begin{cases} 
    2 & (x, y) \in A \\
    0 & \text{otherwise}
  \end{cases}
\]

- This is a uniform distribution on \( A \).
Example (continued)

- In general, if $A \subset \mathbb{R}^n$ is a set with finite, positive area $|A|$, then

$$f(x) = \frac{1}{|A|} 1_A(x)$$

is a PDF of the uniform distribution on $A$.

- We can compute

$$P(X + Y \leq 1/2) = \int \int_{x+y \leq 1/2} f(x, y) \, dx \, dy$$

$$= \int_0^{1/2} \int_0^{1/2-y} 2 \, dx \, dy$$

$$= \int_0^{1/2} \left( \int_0^{1/2-y} 2 \, dx \right) \, dy$$

$$= \int_0^{1/2} \left( 1 - 2y \right) \, dy$$

$$= \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$$
Example (continued)

- In general, if $X$ is uniform on $A$, then
  \[ P(X \in B) = \frac{|A \cap B|}{|A|} \]

- We can compute the marginal CDF of $X$: For $0 \leq a \leq 1$
  \[
  F_X(a) = P(X \leq a) = \int_{-\infty}^{a} \int_{-\infty}^{\infty} f(x, y) dy dx \\
  = \int_{0}^{a} \int_{0}^{1-x} 2 dy dx \\
  = \int_{0}^{a} 2 - 2x dx = 2a - a^2
  \]

- The marginal density of $X$ is
  \[
  f_X(x) = (2 - 2x)1_{[0,1]}(x) = 2(1 - x)1_{[0,1]}(x)
  \]
• In general,

\[
F_X(a) = \int_{-\infty}^{a} \int_{-\infty}^{\infty} f(x, y) \, dy \, dx
\]

\[
f_X(x) = \int_{-\infty}^{\infty} f(x, y) \, dy
\]

• We can compute expectations as integrals:

\[
E[g(X_1, \ldots, X_n)] = \int \cdots \int g(x_1, \ldots, x_n) f(x_1, \ldots, x_n) \, dx_1 \cdots dx_n
\]
• For our example,

\[ E[X + Y] = \int_0^1 \int_0^{1-x} (x + y)^2 \, dy \, dx \]

\[ = \int_0^1 (2xy + y^2)|_{0}^{1-x} \, dx \]

\[ = \int_0^1 2x(1 - x) + (1 - x)^2 \, dx \]

\[ = \int_0^1 (1 + x)(1 - x) \, dx \]

\[ = \int_0^1 (1 - x^2) \, dx = 1 - \frac{1}{3} = \frac{2}{3} \]
• Or we can find

\[
E[X] = \int_0^1 x^2 (1 - x) \, dx = x^2 - \frac{2}{3} x^3 \bigg|_0^1 = 1 - \frac{2}{3} = \frac{1}{3}
\]

\[
E[Y] = \frac{1}{3} \quad \text{by symmetry}
\]

and

\[
E[X + Y] = E[X] + E[Y] = \frac{2}{3}
\]

• A joint density function satisfies

\[
f(x_1, \ldots, x_n) \geq 0
\]

\[
\int \cdots \int f(x_1, \ldots, x_n) \, dx_1 \cdots dx_n = 1
\]

• Any function \( f \) with these properties is a joint density.
It is possible for marginal distributions to be continuous but the joint distribution is not.

For example, suppose $X$ and $Y$ are jointly continuous and $Z = X + Y$.

- All three of $X$, $Y$, and $Z$ are continuous.
- Each of the pairs $(X, Y)$, $(X, Z)$, and $(Y, Z)$ is jointly continuous.
- But $(X, Y, Z)$ is not jointly continuous.
- The distribution of $(X, Y, Z)$ is concentrated on a two-dimensional plane.

It is also possible to have mixtures of

- discrete components
- components continuous on a lower dimensional subspace
- jointly continuous components

For example, let $X$ be the amount of rainfall today, $Y$ the amount of rainfall tomorrow.
We say that \( n \) random variables \( X_1, \ldots, X_n \) are (mutually) independent if for "any" \( A_1, \ldots, A_n \)

\[
P(X_1 \in A_1, \ldots, X_n \in A_n) = P(X \in A_1) \times \cdots \times P(X_n \in A_n)
\]

**Theorem**

*If \( X_1, \ldots, X_n \) are discrete (or jointly continuous) then they are independent if and only if*

\[
f(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n)
\]

*for all \( x_1, \ldots, x_n \).*
Proof.

For $n = 2$ and the discrete case:

- Independence implies factorization:

  $$f(x, y) = P(X = x, Y = y) = P(X = x)P(Y = y) = f_X(x)f_Y(y)$$

- Factorization implies independence:

  $$P(X \in A, Y \in B) = \sum_{x \in A} \sum_{y \in B} f(x, y)$$
  $$= \sum_{x \in A} \sum_{y \in B} f_X(x)f_Y(y)$$
  $$= \left(\sum_{x \in A} f_X(x)\right)\left(\sum_{y \in B} f_Y(y)\right)$$
  $$= P(X \in A)P(Y \in B)$$
Example (continued)

- For our running continuous example,

\[ f_X(x)f_Y(y) = 4(1 - x)(1 - y)1_{[0,1]}(x)1_{[0,1]}(y) \]

- However,

\[ f(x, y) = 2 \times 1_{\{(x, y): x \geq 0, y \geq 0, x+y \leq 1\}}(x, y) \]

- So \( X \) and \( Y \) are not independent.

- If \( X, Y \) is uniform on \([0, 1] \times [0, 1]\) then \( X \) and \( Y \) are independent.
Theorem

If $X, Y$ are independent and $g, h$ are nonnegative or satisfy $E[|g(X)|] < \infty$ and $E[|h(Y)|] < \infty$, then

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)].$$
Proof.

For the discrete case:

\[
E[g(X)h(Y)] = \sum_x \sum_y g(x)h(y)f(x, y)
\]

\[
= \sum_x \sum_y g(x)f_X(x)h(y)f_Y(y)
\]

\[
= \left( \sum_x g(x)f_X(x) \right) \left( \sum_y h(y)f_Y(y) \right)
\]

\[
= E[g(X)]E[h(Y)]
\]
• This holds for any number of independent random variables with the obvious generalization.

• Suppose $X, Y$ are independent and $E[X^2], E[Y^2]$ are finite.

• Then

\[
\text{Var}(X + Y) = E[((X + Y) - E[X] - E[Y])^2] \\
= E[((X - E[X]) + (Y - E[Y]))^2] \\
= E[(X - E[X])^2] + E[(Y - E[Y])^2] \\
\quad + 2E[(X - E[X])(Y - E[Y])] \\
= \text{Var}(X) + \text{Var}(Y) + 0 \\
= \text{Var}(X) + \text{Var}(Y)
\]

• This extends to any number of variables.
**Example**

- Let $Y_1, \ldots, Y_n$ be independent Bernoulli($p$) random variables.
- Then $X = Y_1 + \cdots + Y_n$ is Binomial($n, p$).
- The variance of $X$ is therefore

\[
\text{Var}(X) = \text{Var}(Y_1 + \cdots + Y_n) \\
= \text{Var}(Y_1) + \cdots + \text{Var}(Y_n) \\
= np(1 - p)
\]
Another important consequence of independence:

**Theorem**

Let $X, Y$ be independent with MGF’s $M_X, M_Y$. Then

$$M_{X+Y}(t) = M_X(t)M_Y(t)$$

**Proof.**

$$M_{X+Y}(t) = E[\exp\{t(X + Y)\}]$$

$$= E[\exp\{tX\}\exp\{tY\}]$$

$$= E[e^{tX}]E[e^{tY}]$$

$$= M_X(t)M_Y(t)$$
Example

• If $Y_1, \ldots, Y_n$ are independent Bernoulli($p$), then

$$M_{Y_i}(t) = pe^t + 1 - p.$$ 

• So if $X = Y_1 + \cdots + Y_n$ is binomial, then

$$M_X(t) = M_{Y_1}(t) \times \cdots \times M_{Y_n}(t) = (pe^t + 1 - p)^n$$
Example

- If $Y_1, \ldots, Y_r$ are independent Geometric($p$), then

\[
M_{Y_i} = \begin{cases} 
  \frac{pe^t}{1-(1-p)e^t} & t < -\log(1-p) \\
  \infty & \text{otherwise}
\end{cases}
\]

- So if $X = Y_1 + \cdots + Y_r$, then

\[
M_X(t) = M_{Y_1}(t) \times \cdots \times M_{Y_r}(t) = \begin{cases} 
  \left(\frac{pe^t}{1-(1-p)e^t}\right)^r & t < -\log(1-p) \\
  \infty & \text{otherwise}
\end{cases}
\]

- So $X$ has a negative binomial distribution.
Recap

- Jointly discrete random variables
- Jointly continuous random variables
- Independent random variables
- Independence and expectation
- Variance of sums of independent random variables
- MGF of a sum of independent random variables
Example

- Perhaps the most important example: Suppose $X$ and $Y$ are independent and

  $$X \sim N(\mu_X, \sigma_X^2)$$
  $$Y \sim N(\mu_Y, \sigma_Y^2)$$

- Then

  $$M_{X+Y}(t) = \exp\left\{ t\mu_X + \frac{1}{2} t^2 \sigma_X^2 \right\} \exp\left\{ t\mu_Y + \frac{1}{2} t^2 \sigma_Y^2 \right\}$$

  $$= \exp\left\{ t(\mu_X + \mu_Y) + \frac{1}{2} t^2 (\sigma_X^2 + \sigma_Y^2) \right\}$$

- So $X \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$. 
Covariance and Correlation

• Suppose $X$, $Y$ are not independent.
• Can we somehow quantify “degrees of dependence”? 
• The covariance captures some of this:

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

• The covariance can also be computed as:

$$\text{Cov}(X, Y) = E[XY - X\mu_Y - Y\mu_X + \mu_X \mu_Y]$$

$$= E[XY] - E[X]\mu_Y - E[Y]\mu_X + \mu_X \mu_Y$$

$$= E[XY] - \mu_X \mu_Y$$

• If $X$, $Y$ are independent (and $E[X^2], E[Y^2]$ are finite), then $\text{Cov}(X, Y) = 0$. 
• However, \( X \sim N(0, 1), Y = X^2 \) gives

\[
E[XY] = E[X^3] = 0
\]

• Therefore

\[
\text{Cov}(X, Y) = 0
\]

• But \( X, Y \) are very strongly dependent.

• Covariance captures only \textit{linear} dependence, not quadratic or other dependence.

• The sign of the covariance is meaningful:
  \(+\): \( X, Y \) tend to be on the same side of their means
  \(-\): \( X, Y \) tend to be on opposite sides of their means

• The magnitude is not meaningful—it depends on the units of measurement.
• To get a dimensionless quantity, let

\[ \rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \]

• \( \rho_{X,Y} \) is the correlation coefficient of \( X, Y \).

• We define \( \rho_{X,Y} = 0 \) if \( \sigma_X = 0 \) or \( \sigma_Y = 0 \).

**Theorem**

If \( X, Y \) have \( E[X^2], E[Y^2] \) finite, then \( \rho_{X,Y} \) is defined, and

\[-1 \leq \rho_{X,Y} \leq 1\]

Furthermore, if \( \rho_{X,Y} = \pm 1 \), then for some constants \( a, b, c \) with \( a \neq 0 \) and \( b \neq 0 \), \( aX + bY + c \) is identically zero.
Proof.

- Let
  \[ h(t) = E[((X - \mu_X)t + (Y - \mu_Y))^2]. \]
- Then \( h(t) \geq 0 \) for all \( t \).
- But
  \[ h(t) = t^2\sigma_X^2 + 2t\text{Cov}(X, Y) + \sigma_Y^2. \]
- So
  \[ (2\text{Cov}(X, Y))^2 - 4\sigma_X^2\sigma_Y^2 \leq 0 \]
  or
  \[ |\text{Cov}(X, Y)| \leq \sigma_X\sigma_Y. \]
- Equality holds if and only if for some \( t^* \) we have \( h(t^*) = 0 \), or
  \[ 0 = E[((X - \mu_X)t^* + (Y - \mu_Y))^2] \]
- This implies that \((X - \mu_X)t^* + (Y - \mu_Y) = 0\) almost surely.
Example

• For our simple uniform on a triangle example,

\[
E[XY] = \int_0^1 \int_0^{1-x} xy \, 2dydx
\]

\[
= \int_0^1 x(1-x)^2 \, dx
\]

\[
= \int_0^1 x - 2x^2 + x^3 \, dx
\]

\[
= \frac{1}{2} - \frac{2}{3} + \frac{1}{4} = \frac{3}{4} - \frac{2}{3} = \frac{1}{12}
\]

• The means are \(E[X] = E[Y] = \frac{1}{3}\).

• Therefore

\[
\text{Cov}(X, Y) = \frac{1}{12} - \frac{1}{9} = -\frac{1}{36}
\]
Example (continued)

- The variances are

\[
\text{Var}(X) = \text{Var}(Y) = \int_0^1 x^2 2(1 - x) \, dx - \frac{1}{9}
\]

\[
= \int_0^1 2x^2 - 2x^3 \, dx - \frac{1}{9}
\]

\[
= \frac{2}{3} - \frac{2}{4} - \frac{1}{9} = \frac{24 - 18 - 4}{36} = \frac{2}{36}
\]

- The correlation is therefore

\[
\rho_{X,Y} = -\frac{1}{2}
\]

- Negative correlation makes sense:
**Some Properties of Covariance**

For random variables $X, Y, Z$ with finite variances and a constant $c$

\[
\begin{align*}
\text{Var}(X) &= \text{Cov}(X, X) \\
\text{Cov}(X, Y) &= \text{Cov}(Y, X) \\
\text{Cov}(X + c, Y) &= \text{Cov}(X, Y) \\
\text{Cov}(cX, Y) &= c\text{Cov}(X, Y) \\
\text{Cov}(X + Y, Z) &= \text{Cov}(X, Z) + \text{Cov}(Y, Z)
\end{align*}
\]
Theorem

Let $X_1, \ldots, X_n$ have $E[X_i^2] < \infty$. Then

$$\text{Var}(X_1 + \cdots + X_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} \text{Cov}(X_i, X_j)$$

$$= \sum_{i=1}^{n} \text{Var}(X_i) + 2 \sum_{i<j} \text{Cov}(X_i, X_j)$$
Proof.

\[
\text{Var} \left( \sum X_i \right) = \text{Cov} \left( \sum_{i} X_i, \sum_{j} X_j \right) \\
= \sum_{i} \text{Cov} \left( X_i, \sum_{j} X_j \right) \\
= \sum_{i} \sum_{j} \text{Cov}(X_i, X_j) \\
= \sum_{i=1}^{n} \text{Var}(X_i) + \sum_{i \neq j} \text{Cov}(X_i, X_j) \\
= \sum_{i=1}^{n} \text{Var}(X_i) + 2 \sum_{i < j} \text{Cov}(X_i, X_j)
\]
Example

- Have $N$ items, $M$ defectives.
- Choose $K$ at random, without replacement.
- $X$ is the number of defectives in the sample.
- $X$ has a hypergeometric distribution.
- Let $Y_i = 1$ if $i$-th item chosen is defective, 0 otherwise.
- Then $X = Y_1 + \cdots + Y_K$.
- The $Y_i$ are Bernoulli($p = \frac{M}{N}$), so

$$E[Y_i] = \frac{M}{N}$$

$$Var(Y_i) = \frac{M}{N} \left( 1 - \frac{M}{N} \right)$$
Example (continued)

- To compute $\text{Var}(X)$ we need $\text{Cov}(Y_i, Y_j)$ for $i \neq j$:

  \[
  E[Y_i Y_j] = E[Y_1 Y_2] = P(1\text{st and 2nd defective}) = \frac{M (M - 1)}{N (N - 1)}
  \]

- So

  \[
  \text{Cov}(Y_1, Y_2) = \frac{M (M - 1)}{N (N - 1)} - \left( \frac{N}{N} \right)^2 = \frac{M}{N} \left( \frac{M - 1}{N - 1} - \frac{M}{N} \right) = \frac{M}{N} \left( \frac{MN - N - MN + M}{N(N - 1)} \right) = \frac{M}{N} \left( \frac{M - N}{N} \right) \frac{1}{N - 1} = -\frac{M}{N} \left( 1 - \frac{N}{M} \right) \frac{1}{N - 1}
  \]
Example (continued)

- Thus

\[
\text{Var}(X) = \sum_{i=1}^{K} \text{Var}(Y_i) + \sum_{i \neq j} \text{Cov}(Y_i, Y_j) \\
= K \text{Var}(Y_1) + K(K-1) \text{Cov}(Y_1, Y_2) \\
= K \frac{M}{N} \left(1 - \frac{M}{N}\right) - K(K-1) \frac{M}{N} \left(1 - \frac{M}{N}\right) \frac{1}{N-1} \\
= K \frac{M}{N} \left(1 - \frac{M}{N}\right) \left(1 - \frac{K-1}{N-1}\right) \\
= K \frac{M}{N} \left(1 - \frac{M}{N}\right) \frac{N-K}{N-1}
\]
Example

- In the matching problem with \( n \geq 2 \) hats let \( X \) be the number of matches.
- Let \( Y_i = 1 \) if person \( i \) gets their hat and let \( Y_i = 0 \) otherwise.
- Then \( X = Y_1 + \cdots + Y_n \).
- The \( Y_i \) are Bernoulli with success probability \( 1/n \).
- The expected product for \( i \neq j \) is

\[
E[Y_i Y_j] = E[Y_1 Y_2] = \frac{1}{n} \times \frac{1}{n-1}.
\]

- So

\[
\text{Cov}(Y_i, Y_j) = \frac{1}{n(n-1)} - \frac{1}{n^2} = \frac{1}{n} \left( \frac{1}{n-1} - \frac{1}{n} \right) = \frac{1}{n^2(n-1)}.
\]
Example

- The variance of $X$ is therefore

$$\text{Var}(X) = \sum_{i=1}^{n} \text{Var}(Y_i) + \sum_{i\neq j} \text{Cov}(Y_i, Y_j)$$

$$= n \frac{1}{n} \left( 1 - \frac{1}{n} \right) + n(n-1) \frac{1}{n^2(n-1)}$$

$$= 1.$$

- For $n = 1$ the variance is zero.
Example

- Let $\mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho$ be real numbers with
  - $\sigma_X, \sigma_Y > 0$
  - $-1 < \rho < 1$.

- Define the joint density function

\[
f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y}\sqrt{1-\rho^2} \times \exp\left\{ -\frac{1}{2(1-\rho^2)} \left[ \left( \frac{x-\mu_X}{\sigma_X} \right)^2 + \left( \frac{y-\mu_Y}{\sigma_Y} \right)^2 \right. \right.
\]
\[
\left. \left. -2\rho \left( \frac{x-\mu_X}{\sigma_X} \right) \left( \frac{y-\mu_Y}{\sigma_Y} \right) \right] \right\}
\]

- This is the bivariate normal density.
Bivariate and Multivariate Normal Distribution

\[ \rho = 0.5 \]

\[ \rho = -0.8 \]

\[ \rho = 0.5 \]

\[ \rho = -0.8 \]
Example (continued)

• Basic properties:

\[ X \sim N(\mu_X, \sigma_X^2) \]
\[ Y \sim N(\mu_Y, \sigma_Y^2) \]
\[ \rho_{X,Y} = \rho \]

• Furthermore, for any \( a, b \)

\[ aX + bY \sim N(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\rho\sigma_X\sigma_Y) \]
Recap

- MGF of a sum of independent random variables
- Covariance and correlation
- Variance of a sum of correlated random variables
- Variance of the hypergeometric distribution
- Bivariate and multivariate normal distribution
Example (continued)

Alternate form:

- Let
  \[
  Z = \begin{bmatrix} X \\ Y \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \quad C = \begin{bmatrix} \sigma_X^2 & \rho \sigma_X \sigma_Y \\ \rho \sigma_X \sigma_Y & \sigma_Y^2 \end{bmatrix}
  \]

- Then, for \( n = 2 \),
  \[
  f_Z(z) = \frac{1}{(2\pi)^{n/2}(\det C)^{1/2}} \exp \left\{ -\frac{1}{2} (z - \mu)^T C^{-1} (z - \mu) \right\}.
  \]

- If
  \[
  d = \begin{bmatrix} a \\ b \end{bmatrix}
  \]
  then \( d^T Z = aX + bY \) and
  \[
  d^T Z \sim N(d^T \mu, d^T Cd).
  \]
Example (continued)

• $f_Z$ is a proper density for any $n \geq 1$ if $C$ is
  • symmetric
  • strictly positive definite, i.e.
    \[ d^T Cd > 0 \]
    for all non-zero column vectors $d \in \mathbb{R}^n$.

• $C$ is the variance-covariance matrix, or covariance matrix of $Z$, i.e.
  \[ C_{ij} = \text{Cov}(Z_i, Z_j). \]

• All marginals and joint marginals are univariate or multivariate normal.

• For any non-zero column vector of constants $d$
  \[ d^T Z \sim N(d^T \mu, d^T Cd). \]
Suppose $X_1, \ldots, X_n$ are random variables and $a_1, \ldots, a_n$ are constants.

We are often interested in linear combinations of the form

$$\sum_{i=1}^{n} a_i X_i.$$ 

Some examples are

- the sum
  $$Y = \sum_{i=1}^{n} X_i$$

- the average
  $$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i = \frac{1}{n} Y$$

- the residuals
  $$X_i - \bar{X}.$$
• Suppose the $X_i$ have finite variances and let

$$\mu_i = E[X_i]$$

$$\sigma^2_i = \text{Var}(X_i)$$

$$\sigma_{ij} = \text{Cov}(X_i, X_j)$$

• The mean of a linear combination is

$$E \left[ \sum_{i=1}^{n} a_i X_i \right] = \sum_{i=1}^{n} E[a_i X_i] = \sum_{i=1}^{n} a_i E[X_i] = \sum_{i=1}^{n} a_i \mu_i$$

• The variance of a linear combination is

$$\text{Var} \left( \sum_{i=1}^{n} a_i X_i \right) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \text{Cov}(X_i, X_j) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \sigma_{ij}$$
• These expressions can be simplified using matrix notation.
• Let

\[ X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}, \quad a = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}, \quad C = \begin{bmatrix} \sigma_{11} & \ldots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \ldots & \sigma_{nn} \end{bmatrix} \]

• Then

\[ a^T X = \sum_{i=1}^{n} a_i X_i \]

\[ E[a^T X] = a^T \mu \]

\[ \text{Var}(a^T X) = a^T Ca \]

• If \( b^T X \) is another linear combination, then

\[ \text{Cov}(a^T X, b^T X) = a^T C b \]
• Suppose $A$ is an $m \times n$ matrix of constants.
• Then $Y = AX$ is a vector of $m$ linear combinations of the elements of $X$.
• The vector of means of $Y$ is
  \[
  E[Y] = \begin{bmatrix}
  E[Y_1] \\
  \vdots \\
  E[Y_m]
  \end{bmatrix} = A\mu
  \]
• The covariance matrix of $Y$ is
  \[AC\mathbf{A}^T\]
• These results can be used, for example, to derive the covariance matrix of least squares estimators in linear regression.
Copulas

- Suppose $Z_1, Z_2$ are bivariate normal with means $\mu_1 = \mu_2 = 0$, variances $\sigma_1^2 = \sigma_2^2 = 1$, and correlation $\rho$.
- Let $U_i = \Phi(Z_i)$, where $\Phi$ is the standard normal CDF.
- Then $U_1, U_2$ have uniform $[0, 1]$ marginal distributions.
- If $F_1$ and $F_2$ are two CDFs define variables $Y_i = F_i^{-1}(U_i)$.
- The resulting variables have marginal CDFs $Y_i \sim F_i$ and are dependent if $\rho \neq 0$.
- A distribution on the $n$-dimensional unit cube with uniform one-dimensional marginals is called a copula.
- The distribution of $U_1, U_2$ is called a Gaussian copula.
- Other forms of copulas are available.
- Most incorporate a small number of dependence parameters.
- Copulas are a useful mechanism for creating models for dependent variables.
Suppose $X, Y$ are discrete.

**Definition**

The *conditional distribution* of $X$ given $Y = y$ is the discrete distribution with probability mass function

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)} = P(X = x|Y = y)$$

for $y$ with $P(Y = y) = f_Y(y) > 0$, and is undefined if $f_Y(y) = 0$. 
Example

Uniform distribution on integer triangle (example from before):

\[ f_{X|Y}(x) = \begin{cases} 
 1/3 & \text{if } y = 0, \\
 1/2 & \text{if } y = 1, \\
 1/2 & \text{if } y = 2. 
\end{cases} \]

\[ y = 0: \quad \begin{array}{ccc}
  x & 0 & 1 & 2 \\
  f_{X|Y} & 1/3 & 1/3 & 1/3 \\
\end{array} \]

\[ y = 1: \quad \begin{array}{cc}
  x & 0 & 1 \\
  f_{X|Y} & 1/2 & 1/2 \\
\end{array} \]

\[ y = 2: \quad X \equiv 0. \]
• The conditional PMF can be used to compute *conditional expectations*:

\[ E[g(X)|Y = y] = \sum g(x)f_{X|Y}(x|y) \]

or conditional variances.

• For our example:

\[
\begin{align*}
E[X|Y = 0] &= 1 \\
E[X|Y = 1] &= \frac{1}{2} \\
E[X|Y = 0] &= 0
\end{align*}
\]
• When $Y$ is continuous we have a problem:

$$P(X \in A | Y = y) = \frac{P(X \in A, Y = y)}{P(Y = y)} = \frac{0}{0} = \text{???}$$

• In the jointly continuous case we can proceed by analogy to the discrete case and define the conditional density of $X$ given $Y = y$ as

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}$$

if $f_Y(y) > 0$ and undefined otherwise.

• Again this is a proper density and can be used to find means, variances, etc.
Example

- For our simple uniform distribution on a triangle we have

\[
f(x, y) = \begin{cases} 
2 & x, y \geq 0, x + y \leq 1 \\
0 & \text{otherwise}
\end{cases}
\]

\[
f_Y(y) = \begin{cases} 
2(1 - y) & 0 \leq y \leq 1 \\
0 & \text{otherwise}
\end{cases}
\]

- For \(0 < y < 1\) we get

\[
f_{X|Y}(x|y) = \begin{cases} 
\frac{1}{1-y} & 0 \leq x \leq 1 - y \\
0 & \text{otherwise}
\end{cases}
\]

- This is a uniform distribution on \([0, 1 - y]\).
Example (continued)

- This is a uniform distribution on $[0, 1 - y]$.
- So the conditional mean and variance are

$$E[X|Y = y] = \frac{1 - y}{2}$$

$$\text{Var}(X|Y = y) = \frac{(1 - y)^2}{12}$$
The conditional CDF of $X$ given a continuous $Y$ can be viewed as

$$F_{X|Y}(x|y) = \lim_{h \to 0} P(X \leq x | y - h \leq Y \leq y + h)$$

There are some risks to this view.

But this view is useful when used carefully.

Problem 4.61 and the Miscellanea section at the end of Chapter 4 address this issue.
There is an analog for expectations to the law of total probability:

\[ E[X] = \int E[X|Y = y]f_Y(y) \, dy \]

\[ E[X] = \sum E[X|Y = y]f_Y(y) \]

This is sometimes called the *law of total expectation*. 
Proof.

Fort the discrete case:

\[
E[X] = \sum_y \sum_x xf(x, y)
\]

\[
= \sum_y \sum_x x \frac{f(x, y)}{f_Y(y)} f_Y(y)
\]

\[
= \sum_y \sum_x xf_{X|Y}(x|y) f_Y(y)
\]

\[
= \sum_Y E[X|Y = y] f_Y(y)
\]
• To allow a more compact statement, let

\[ h(y) = E[X | Y = y] \]

• So \( h \) is a function of \( y \).
• Let \( E[X | Y] = h(Y) \).
• Then \( E[X | Y] = h(Y) \) is a random variable.
• The law of total expectation becomes

\[ E[X] = E[E[X | Y]] \]

• This holds in the discrete and jointly continuous cases, and all other cases.
Example

- The number $N$ of customers who shop at a store on a particular day has a Poisson distribution with mean $\lambda$.
- The amount spent by customer $i$ is a random variable $X_i$ with mean $\mu$.
- The $X_i$ are mutually independent and independent of $N$.
- We would like to compute the expected total amount spent by all customers on the day in question,

$$S = \sum_{i=1}^{N} X_i$$
Example (continued)

- The conditional expectation of $S$ given $N = n$ is

\[
E[S|N = n] = E \left[ \sum_{i=1}^{N} X_i \bigg| N = n \right]
\]

\[
= E \left[ \sum_{i=1}^{n} X_i \bigg| N = n \right] \quad \text{fix upper limit}
\]

\[
= E \left[ \sum_{i=1}^{n} X_i \right] \quad \text{use independence}
\]

\[
= n \mu
\]

- The conditional expectation of $S$ given $N$ is the random variable

\[
E[S|N] = N \mu
\]

- The mean of $S$ is therefore

\[
E[S] = E[E[S|N]] = E[N \mu] = E[N] \mu = \lambda \mu
\]