Recap: Conditional Expectation

- The conditional expectation of $h(X)$ conditioned on Y is defined by

$$E[h(X)|Y = y] = \sum_x h(x)\, f_{X|Y}(x|y), \quad X, Y \text{ are discrete}$$

$$E[h(X)|Y = y] = \int_{-\infty}^{\infty} h(x)\, f_{X|Y}(x|y)\, dx, \quad X, Y \text{ have joint density}$$

- The conditional expectation of $h(X)$ conditioned on Y is a function of Y:
  $E[h(X)|Y] = g(Y)$; the above specify the value of $g(y)$.
- We define $E[h(X,Y)|Y]$ also as above:

$$E[h(X,Y)|Y = y] = \int_{-\infty}^{\infty} h(x,y)\, f_{X|Y}(x|y)\, dx$$

- If X, Y are independent, $E[h(X)|Y] = E[h(X)]$

Recap: Properties of Conditional Expectation

- It has all the properties of expectation:
  - $E[a|Y] = a$ where a is a constant
  - $E[ah_1(X) + bh_2(X)|Y] = aE[h_1(X)|Y] + bE[h_2(X)|Y]$
  - $h_1(X) \geq h_2(X) \;\; \Rightarrow \;\; E[h_1(X)|Y] \geq E[h_2(X)|Y]$
- Conditional expectation also has some extra properties which are very important
  - $E\left[E[h(X)|Y]\right] = E[h(X)]$
  - $E[h_1(X)h_2(Y)|Y] = h_2(Y)E[h_1(X)|Y]$
  - $E[h(X,Y)|Y = y] = E[h(X,y)|Y = y]$

Recap

- The property of conditional expectation

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

is very useful in calculating expectations

$$EX = \sum_y E[X|Y=y]\ f_Y(y) \quad \text{or}$$

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

- We saw many examples.

Sum of random number of random variables

- Let $X_1, X_2, \cdots$ be iid rv on the same probability space. Suppose $EX_i = \mu < \infty,\ \forall i$.
- Let N be a positive integer valued rv that is independent of all X_i ($EN < \infty$)
- Let $S = \sum_{i=1}^{N} X_i$.
- We want to calculate ES.
- We can use

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

- We have

$$
\begin{aligned}
E[S|N=n] &= E\left[\sum_{i=1}^{N} X_i \mid N=n\right] \\
&= E\left[\sum_{i=1}^{n} X_i \mid N=n\right] \\
&\quad \text{since } E[h(X,Y)|Y=y] = E[h(X,y)|Y=y] \\
&= \sum_{i=1}^{n} E[X_i \mid N=n] = \sum_{i=1}^{n} E[X_i] = n\mu
\end{aligned}
$$

- Hence we get

$$
E[S|N] = N\mu \quad \Rightarrow \quad E[S] = E[N]E[X_1]
$$

Wald's formula

- We took $S = \sum_{i=1}^{N} X_i$ with N independent of all X_i.
- With iid X_i, the formula $ES = EN\ EX_1$ is valid even under some dependence between N and X_i.
- Here are one version of assumptions needed.
  - A1 $E[|X_1|] < \infty$ and $EN < \infty$ (X_i iid).
  - A2 $E\left[X_n\ I_{[N \geq n]}\right] = E[X_n]P[N \geq n],\ \forall n$
- Let $S_N = \sum_{i=1}^{N} X_i$.
- Then, $ES_N = EX_1\ EN$
- Suppose the event $[N \leq n-1]$ depends only on $X_1, \cdots, X_{n-1}$.
- Such an N is called a stopping time.
- Then the event $[N \leq n-1]$ and hence its complement $[N \geq n]$ is independent of X_n and hence $A2$ holds.

Wald's formula

- In the general case, we do not need X_i to be iid.
- Here is one version of this Wald's formula. We assume
  1. $E[|X_i|] < \infty, \ \forall i$ and $EN < \infty$.
  2. $E\left[X_n \ I_{[N \geq n]}\right] = E[X_n]P[N \geq n], \ \forall n$
- Let $S_N = \sum_{i=1}^{N} X_i$ and let $T_N = \sum_{i=1}^{N} E[X_i]$.
- Then, $ES_N = ET_N$.
  If $E[X_i]$ is same for all i, $ES_N = EX_1 \ EN$.

Variance of random sum

- $S = \sum_{i=1}^{N} X_i$, X_i iid, ind of N. Want $\text{Var}(S)$

$$E[S^2] = E\left[\left(\sum_{i=1}^{N} X_i\right)^2\right] = E\left[E\left[\left(\sum_{i=1}^{N} X_i\right)^2 \mid N\right]\right]$$

- As earlier, we have

$$\begin{aligned} E\left[\left(\sum_{i=1}^{N} X_i\right)^2 \mid N = n\right] &= E\left[\left(\sum_{i=1}^{n} X_i\right)^2 \mid N = n\right] \\ &= E\left[\left(\sum_{i=1}^{n} X_i\right)^2\right] \end{aligned}$$

- Let $Y = \sum_{i=1}^{n} X_i, \quad X_i$ iid
- Then, $\text{Var}(Y) = n \; \text{Var}(X_1)$
- Hence we have

$$E[Y^2] = \text{Var}(Y) + (EY)^2 = n \; \text{Var}(X_1) + (nEX_1)^2$$

- Using this

$$E\left[\left(\sum_{i=1}^{N} X_i\right)^2 \mid N = n\right] = E\left[\left(\sum_{i=1}^{n} X_i\right)^2\right] = n \; \text{Var}(X_1) + (nEX_1)^2$$

- Hence

$$E\left[\left(\sum_{i=1}^{N} X_i\right)^2 \mid N\right] = N \; \text{Var}(X_1) + N^2 \; (EX_1)^2$$

- $S = \sum_{i=1}^{N} X_i \quad (X_i$ iid). We got

$$E[S^2] = E[\ E[S^2|N]\ ] = EN\ \mathsf{Var}(X_1) + E[N^2](EX_1)^2$$

- Now we can calculate variance of S as

$$\begin{aligned}\mathsf{Var}(S) &= E[S^2] - (ES)^2 \\ &= EN\ \mathsf{Var}(X_1) + E[N^2](EX_1)^2 - (EN\ EX_1)^2 \\ &= EN\ \mathsf{Var}(X_1) + (EX_1)^2 \left(E[N^2] - (EN)^2\right) \\ &= EN\ \mathsf{Var}(X_1) + \mathsf{Var}(N)\ (EX_1)^2\end{aligned}$$

Another Example

- We toss a (biased) coin till we get k consecutive heads. Let N_k denote the number of tosses needed.
- N_1 would be geometric.
- We want $E[N_k]$. What rv should we condition on?
- Useful rv here is N_{k-1}

$$E[N_k \mid N_{k-1} = n] = (n+1)p + (1-p)(n+1+E[N_k])$$

- Thus we get the recurrence relation

$$\begin{aligned} E[N_k] &= E[\ E[N_k \mid N_{k-1}]\ ] \\ &= E\left[(N_{k-1}+1)p + (1-p)(N_{k-1}+1+E[N_k])\right] \end{aligned}$$

- We have

$$E[N_k] = E\left[(N_{k-1}+1)p + (1-p)(N_{k-1}+1+E[N_k])\right]$$

- Denoting $M_k = E[N_k]$, we get

$$\begin{aligned}
M_k &= pM_{k-1} + p + (1-p)M_{k-1} + (1-p) + (1-p)M_k \\
pM_k &= M_{k-1} + 1 \\
M_k &= \frac{1}{p}\,M_{k-1} + \frac{1}{p} \\
&= \frac{1}{p}\left(\frac{1}{p}\,M_{k-2} + \frac{1}{p}\right) + \frac{1}{p} = \left(\frac{1}{p}\right)^2 M_{k-2} + \left(\frac{1}{p}\right)^2 + \frac{1}{p} \\
&= \left(\frac{1}{p}\right)^{k-1} M_1 + \sum_{j=1}^{k-1}\left(\frac{1}{p}\right)^j \\
&= \frac{1-p^k}{(1-p)p^k} \quad \text{taking} \quad M_1 = \frac{1}{p}
\end{aligned}$$

- As mentioned earlier, we can use the conditional expectation to calculate probabilities of events also.

$$P(A) = E[I_A] = E\,[\ E\,[I_A|Y]\ ]$$

$$E[I_A|Y = y] = P[I_A = 1|Y = y] = P(A|Y = y)$$

- Thus, we get

$$\begin{aligned} P(A) &= E[I_A] = E\,[\ E\,[I_A|Y]\ ] \\ &= \sum_y P(A|Y = y)P[Y = y], \quad \text{when } Y \text{ is discrete} \\ &= \int P(A|Y = y)\ f_Y(y)\ dy, \quad \text{when } Y \text{ is continuous} \end{aligned}$$

Example

- Let X, Y be independent continuous rv
- We want to calculate $P[X \leq Y]$
- We can calculate it by integrating joint density over $A = \{(x, y) \ : \ x \leq y\}$

$$
\begin{aligned}
P[X \leq Y] &= \int \int_A f_X(x) \ f_Y(y) \ dx \ dy \\
&= \int_{-\infty}^{\infty} f_Y(y) \ \left(\int_{-\infty}^{y} f_X(x) \ dx \ \right) dy \\
&= \int_{-\infty}^{\infty} F_X(y) \ f_Y(y) \ dy
\end{aligned}
$$

- IF X, Y are *iid* then $P[X < Y] = 0.5$

- We can also use the conditional expectation method here

$$
\begin{aligned}
P[X \leq Y] &= \int_{-\infty}^{\infty} P[X \leq Y \mid Y = y] \; f_Y(y) \; dy \\
&= \int_{-\infty}^{\infty} P[X \leq y \mid Y = y] \; f_Y(y) \; dy \\
&= \int_{-\infty}^{\infty} P[X \leq y] \; f_Y(y) \; dy \\
&= \int_{-\infty}^{\infty} F_X(y) \; f_Y(y) \; dy
\end{aligned}
$$

Another Example

- Consider a sequence of bernoullli trials where p, probability of success, is random.
- We first choose p uniformly over $(0,1)$ and then perform n tosses.
- Let X be the number of heads.
- Conditioned on knowledge of p, we know distribution of X
$$P[X=k \mid p] = {}^{n}C_k \, p^k \, (1-p)^{n-k}$$
- Now we can calculate $P[X=k]$ using the conditioning argument.

- Assuming p is chosen uniformly from $(0,\ 1)$, we get

$$
\begin{aligned}
P[X=k] &= \int [P[X=k \mid p]\ f(p)\ dp \\
&= \int_0^1 {}^nC_k\ p^k\ (1-p)^{n-k}\ 1\ dp \\
&= {}^nC_k\ \frac{k!(n-k)!}{(n+1)!} \\
\text{because} &\int_0^1 p^k\ (1-p)^{n-k}\ dp = \frac{\Gamma(k+1)\Gamma(n-k+1)}{\Gamma(n+2)} \\
&= \frac{1}{n+1}
\end{aligned}
$$

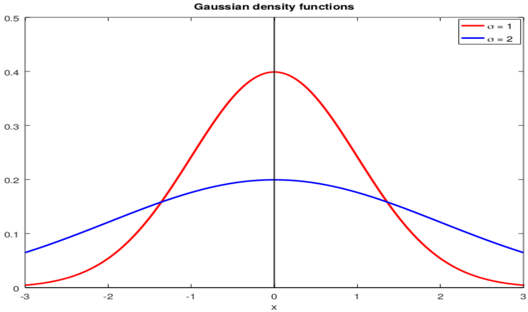
- So, we get: $P[X=k] = \frac{1}{n+1},\ k = 0, 1, \cdots, n$

Gaussian or Normal distribution

- The Gaussian or normal density is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

- If X has this density, we denote it as $X \sim \mathcal{N}(\mu, \sigma^2)$. We showed $EX = \mu$ and $\text{Var}(X) = \sigma^2$
- The density is a 'bell-shaped' curve



- Standard Normal rv — $X \sim \mathcal{N}(0,1)$
- The distribution function of standard normal is

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\ dt$$

- Suppose $X \sim \mathcal{N}(\mu, \sigma^2)$

$$\begin{aligned} P[a \le X \le b] &= \int_{a}^{b} \frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{(x-\mu)^2}{2\sigma^2}}\ dx \\ &\quad \text{take} \ \ y = \frac{(x-\mu)}{\sigma} \ \ \Rightarrow \ \ dy = \frac{1}{\sigma} dx \\ &= \int_{\frac{(a-\mu)}{\sigma}}^{\frac{(b-\mu)}{\sigma}} \frac{1}{\sqrt{2\pi}}\ e^{-\frac{y^2}{2}}\ dy \\ &= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \end{aligned}$$

- We can express probability of events involving all Normal rv using Φ.

- $X \sim \mathcal{N}(0,1)$. Then its mgf is

$$\begin{aligned} M_X(t) &= E\left[e^{tX}\right] = \int_{-\infty}^{\infty} e^{tx} \, \frac{1}{\sqrt{2\pi}} \, e^{-\frac{x^2}{2}} \, dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2-2tx)} \, dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left((x-t)^2-t^2\right)} \, dx \\ &= e^{\frac{1}{2}t^2} \, \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-t)^2} \, dx \\ &= e^{\frac{1}{2}t^2} \end{aligned}$$

- Now let $Y = \sigma X + \mu$. Then $Y \sim \mathcal{N}(\mu, \sigma^2)$.
  The mgf of Y is

$$\begin{aligned} M_Y(t) &= E\left[e^{t(\sigma X+\mu)}\right] = e^{t\mu} \, E\left[e^{(t\sigma)X}\right] = e^{t\mu} \, M_X(t\sigma) \\ &= e^{\left(\mu t + \frac{1}{2}t^2\sigma^2\right)} \end{aligned}$$

Multi-dimensional Gaussian Distribution

- The n-dimensional Gaussian density is given by

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{|\Sigma|^{\frac{1}{2}} \, (2\pi)^{\frac{n}{2}}} \, e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu})}, \ \ \mathbf{x} \in \Re^n$$

- $\boldsymbol{\mu} \in \Re^n$ and $\Sigma \in \Re^{n \times n}$ are parameters of the density and Σ is symmetric and positive definite.
- If $X_1, \cdots, X_n$ have the above joint density, they are said to be jointly Gaussian.
- We denote this by $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$
- We will now show that this is a joint density function.

- We begin by showing the following is a density (when M is symmetric +ve definite)

$$f_{\mathbf{Y}}(\mathbf{y}) = C\ e^{-\frac{1}{2}\mathbf{y}^T M \mathbf{y}}$$

- Let $I = \int_{\Re^n} C\ e^{-\frac{1}{2}\mathbf{y}^T M \mathbf{y}}\ d\mathbf{y}$
- Since M is real symmetric, there exists an orthogonal transform, L with $L^{-1} = L^T$, $|L| = 1$ and $L^T M L$ is diagonal
- Let $L^T M L = \text{diag}(m_1, \cdots, m_n)$.
- Then for any $\mathbf{z} \in \Re^n$,

$$\mathbf{z}^T L^T M L \mathbf{z} = \sum_i m_i z_i^2$$

- We now get

$$
\begin{aligned}
I &= \int_{\Re^n} C\ e^{-\frac{1}{2}\mathbf{y}^T M \mathbf{y}}\ d\mathbf{y} \\
&\quad \text{change variable: } \mathbf{z} = L^{-1}\mathbf{y} = L^T\mathbf{y} \ \ \Rightarrow \mathbf{y} = L\mathbf{z} \\
&= C \int_{\Re^n} e^{-\frac{1}{2}\mathbf{z}^T L^T M L \mathbf{z}}\ d\mathbf{z} \quad (\text{note that} \quad |L| = 1) \\
&= C \int_{\Re^n} e^{-\frac{1}{2}\sum_i m_i z_i^2}\ d\mathbf{z} \\
&= C \prod_{i=1}^{n} \int_{\Re} e^{-\frac{1}{2} m_i z_i^2}\ dz_i = C \prod_{i=1}^{n} \int_{\Re} e^{-\frac{1}{2}\frac{z_i^2}{\frac{1}{m_i}}}\ dz_i \\
&= C \prod_{i=1}^{n} \sqrt{2\pi\ \frac{1}{m_i}}
\end{aligned}
$$

- We will first relate $m_1 \cdots m_n$ to the matrix M.
- By definition, $L^T M L = \text{diag}(m_1, \cdots, m_n)$. Hence

$$\text{diag}\left(\frac{1}{m_1}, \cdots, \frac{1}{m_n}\right) = \left(L^T M L\right)^{-1} = L^{-1} M^{-1} (L^T)^{-1} = L^T M^{-1} L$$

- Since $|L| = 1$, we get

$$\left|L^T M^{-1} L\right| = \left|M^{-1}\right| = \frac{1}{m_1 \cdots m_n}$$

Putting all this together

$$\int_{\Re^n} C \, e^{-\frac{1}{2}\mathbf{y}^T M \mathbf{y}} \, d\mathbf{y} = C \prod_{i=1}^{n} \sqrt{2\pi \, \frac{1}{m_i}} = C \, (2\pi)^{\frac{n}{2}} \left|M^{-1}\right|^{\frac{1}{2}}$$

$$\Rightarrow \frac{1}{(2\pi)^{\frac{n}{2}} \left|M^{-1}\right|^{\frac{1}{2}}} \int_{\Re^n} e^{-\frac{1}{2}\mathbf{y}^T M \mathbf{y}} \, d\mathbf{y} = 1$$

- We showed the following is a density (taking $M^{-1} = \Sigma$)

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{(2\pi)^{\frac{n}{2}} \, |\Sigma|^{\frac{1}{2}}} \, e^{-\frac{1}{2}\mathbf{y}^T \Sigma^{-1} \mathbf{y}}, \quad \mathbf{y} \in \Re^n$$

- Let $\mathbf{X} = \mathbf{Y} + \boldsymbol{\mu}$. Then

$$f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{Y}}(\mathbf{x} - \boldsymbol{\mu}) = \frac{1}{(2\pi)^{\frac{n}{2}} \, |\Sigma|^{\frac{1}{2}}} \, e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu})}$$

- This is the multidimensional Gaussian distribution

- Consider $\mathbf{Y}$ with joint density

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{(2\pi)^{\frac{n}{2}} \, |\Sigma|^{\frac{1}{2}}} \, e^{-\frac{1}{2}\mathbf{y}^T \Sigma^{-1} \mathbf{y}}, \;\; \mathbf{y} \in \Re^n$$

- As earlier let $M = \Sigma^{-1}$. Let $L^T M L = \text{diag}(m_1, \cdots, m_n)$
- Define $\mathbf{Z} = (Z_1, \cdots, Z_n)^T = L^T \mathbf{Y}$. Then $\mathbf{Y} = L\mathbf{Z}$.
- Recall $|L| = 1$, $|M^{-1}| = (m_1 \cdots m_n)^{-1}$
- Then density of $\mathbf{Z}$ is

$$\begin{aligned} f_{\mathbf{Z}}(\mathbf{z}) &= \frac{1}{(2\pi)^{\frac{n}{2}} \, |M^{-1}|^{\frac{1}{2}}} \, e^{-\frac{1}{2}\mathbf{z}^T L^T M L \mathbf{z}} = \frac{1}{(2\pi)^{\frac{n}{2}} \left(\frac{1}{m_1 \cdots m_n}\right)^{\frac{1}{2}}} e^{-\frac{1}{2}\sum_i m_i z_i^2} \\ &= \prod_{i=1}^{n} \sqrt{\frac{1}{2\pi}} \, \frac{1}{\sqrt{\frac{1}{m_i}}} \, e^{-\frac{1}{2} m_i z_i^2} = \prod_{i=1}^{n} \sqrt{\frac{1}{2\pi}} \, \frac{1}{\sqrt{\frac{1}{m_i}}} \, e^{-\frac{1}{2} \frac{z_i^2}{\frac{1}{m_i}}} \end{aligned}$$

This shows that $Z_i \sim \mathcal{N}(0, \frac{1}{m_i})$ and Z_i are independent.

- If $\mathbf{Y}$ has density $f_{\mathbf{Y}}$ and $\mathbf{Z} = L^T Y$ then $Z_i \sim \mathcal{N}(0, \frac{1}{m_i})$ and Z_i are independent. Hence,

$$\Sigma_Z = \text{diag}\left(\frac{1}{m_1}, \cdots, \frac{1}{m_n}\right) = L^T M^{-1} L$$

- Also, since $E[Z_i] = 0$, $\Sigma_Z = E[\mathbf{Z}\mathbf{Z}^T]$.
- Since $\mathbf{Y} = L\mathbf{Z}$, $E[\mathbf{Y}] = 0$ and

$$\Sigma_Y = E[\mathbf{Y}\mathbf{Y}^T] = E[L\mathbf{Z}\mathbf{Z}^T L^T] = LE[\mathbf{Z}\mathbf{Z}^T]L^T = L(L^T M^{-1} L)L^T = M^{-1}$$

- Thus, if $\mathbf{Y}$ has density

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{(2\pi)^{\frac{n}{2}} \, |\Sigma|^{\frac{1}{2}}} \, e^{-\frac{1}{2}\mathbf{y}^T \Sigma^{-1} \mathbf{y}}, \quad \mathbf{y} \in \Re^n$$

then $E\mathbf{Y} = 0$ and $\Sigma_Y = M^{-1} = \Sigma$

- Let $\mathbf{Y}$ have density

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{(2\pi)^{\frac{n}{2}} \, |\Sigma|^{\frac{1}{2}}} \; e^{-\frac{1}{2}\mathbf{y}^T \Sigma^{-1} \mathbf{y}}, \quad \mathbf{y} \in \Re^n$$

- Let $\mathbf{X} = \mathbf{Y} + \boldsymbol{\mu}$. Then

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} \, |\Sigma|^{\frac{1}{2}}} \; e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu})}$$

- We have

$$E\mathbf{X} = E[\mathbf{Y} + \boldsymbol{\mu}] = \boldsymbol{\mu}$$

$$\Sigma_X = E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T] = E[\mathbf{Y}\mathbf{Y}^T] = \Sigma$$

Multi-dimensional Gaussian density

- $\mathbf{X} = (X_1, \cdots, X_n)^T$ are said to be jointly Gaussian if

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} \, |\Sigma|^{\frac{1}{2}}} \, e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu})}$$

- $E\mathbf{X} = \boldsymbol{\mu}$ and $\Sigma_X = \Sigma$.
- Suppose Cov$(X_i, X_j) = 0, \forall i \neq j \Rightarrow \Sigma_{ij} = 0, \forall i \neq j$.
- Then Σ is diagonal. Let $\Sigma =$ diag$(\sigma_1^2, \cdots, \sigma_n^2)$.

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} \; \sigma_1 \cdots \sigma_n} e^{-\frac{1}{2} \sum_{i=1}^{n} \left(\frac{x_i - \mu_i}{\sigma_i}\right)^2} = \prod_{i=1}^{n} \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_i - \mu_i}{\sigma_i}\right)^2}$$

- This implies X_i are independent.
- If $X_1, \cdots, X_n$ are jointly Gaussian then uncorrelatedness implies independence.

- Let $\mathbf{X} = (X_1, \cdots, X_n)^T$ be jointly Gaussian:

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} \, |\Sigma|^{\frac{1}{2}}} \; e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu})}$$

- Let $\mathbf{Y} = \mathbf{X} - \boldsymbol{\mu}$.
- Let $M = \Sigma^{-1}$ and L be such that $L^T M L = \text{diag}(m_1, \cdots, m_n)$
- Let $\mathbf{Z} = (Z_1, \cdots, Z_n)^T = L^T Y$.
- Then we saw that $Z_i \sim \mathcal{N}(0, \frac{1}{m_i})$ and Z_i are independent.
- If $X_1, \cdots, X_n$ are jointly Gaussian then there is a 'linear' transform that transforms them into independent random variables.

Moment generating function

- Let $\mathbf{X} = (X_1, \cdots, X_n)^T$ be jointly Gaussian
- Let $\mathbf{Y} = \mathbf{X} - \boldsymbol{\mu}$ and $\mathbf{Z} = (Z_1, \cdots, Z_n)^T = L^T Y$ as earlier
- The moment generating function of $\mathbf{X}$ is given by

$$
\begin{aligned}
M_{\mathbf{X}}(\mathbf{s}) &= E\left[e^{\mathbf{s}^T \mathbf{X}}\right] \\
&= E\left[e^{\mathbf{s}^T (\mathbf{Y} + \boldsymbol{\mu})}\right] = e^{\mathbf{s}^T \boldsymbol{\mu}} \, E\left[e^{\mathbf{s}^T \mathbf{Y}}\right] \\
&= e^{\mathbf{s}^T \boldsymbol{\mu}} \, E\left[e^{\mathbf{s}^T L \mathbf{Z}}\right] \\
&= e^{\mathbf{s}^T \boldsymbol{\mu}} \, E\left[e^{\mathbf{u}^T \mathbf{Z}}\right] \\
& \qquad \text{where} \quad \mathbf{u} = L^T \mathbf{s} \\
&= e^{\mathbf{s}^T \boldsymbol{\mu}} \, M_{\mathbf{Z}}(\mathbf{u})
\end{aligned}
$$

- Since Z_i are independent, easy to get $M_{\mathbf{Z}}$.
- We know $Z_i \sim \mathcal{N}(0, \frac{1}{m_i})$. Hence

$$M_{Z_i}(u_i) = e^{\frac{1}{2}\frac{1}{m_i}u_i^2} = e^{\frac{u_i^2}{2m_i}}$$

$$M_{\mathbf{Z}}(\mathbf{u}) = E\left[e^{\mathbf{u}^T\mathbf{Z}}\right] = \prod_{i=1}^{n} E\left[e^{u_i Z_i}\right] = \prod_{i=1}^{n} e^{\frac{u_i^2}{2m_i}} = e^{\sum_i \frac{u_i^2}{2m_i}}$$

- We derived earlier

$$M_{\mathbf{X}}(\mathbf{s}) = e^{\mathbf{s}^T\boldsymbol{\mu}}\, M_{\mathbf{Z}}(\mathbf{u}), \quad \text{where} \quad \mathbf{u} = L^T\mathbf{s}$$

- We got

$$M_{\mathbf{X}}(\mathbf{s}) = e^{\mathbf{s}^T \boldsymbol{\mu}} \; M_{\mathbf{Z}}(\mathbf{u}); \;\; \mathbf{u} = L^T \mathbf{s}; \;\; M_{\mathbf{Z}}(\mathbf{u}) = e^{\sum_i \frac{u_i^2}{2m_i}}$$

- Earlier we have shown $L^T M^{-1} L = \text{diag}(\frac{1}{m_1}, \cdots, \frac{1}{m_n})$ where $M^{-1} = \Sigma$. Now we get

$$\frac{1}{2} \sum_i \frac{u_i^2}{m_i} = \frac{1}{2} \mathbf{u}^T (L^T M^{-1} L) \mathbf{u} = \frac{1}{2} \mathbf{s}^T M^{-1} \mathbf{s} = \frac{1}{2} \mathbf{s}^T \Sigma \mathbf{s}$$

- Hence we get

$$M_{\mathbf{X}}(\mathbf{s}) = e^{\mathbf{s}^T \boldsymbol{\mu} \, + \, \frac{1}{2} \mathbf{s}^T \Sigma \mathbf{s}}$$

- This is the moment generating function of multi-dimensional Normal density

- Let X, Y be jointly Gaussian. For simplicity let $EX = EY = 0$.
- Let $\text{Var}(X) = \sigma_x^2$, $\text{Var}(Y) = \sigma_y^2$; let $\rho_{XY} = \rho \quad \Rightarrow \quad \text{Cov}(X, Y) = \rho\sigma_x\sigma_y$.
- Now, the covariance matrix and its inverse are given by

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}; \quad \Sigma^{-1} = \frac{1}{\sigma_x^2\sigma_y^2(1-\rho^2)} \begin{bmatrix} \sigma_y^2 & -\rho\sigma_x\sigma_y \\ -\rho\sigma_x\sigma_y & \sigma_x^2 \end{bmatrix}$$

- The joint density of X, Y is given by

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \, e^{-\frac{1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} - \frac{2\rho xy}{\sigma_x\sigma_y}\right)}$$

- This is the bivariate Gaussian density

- Suppose X, Y are jointly Gaussian (with the density above)
- Then, all the marginals and conditionals would be Gaussian.
- $X \sim \mathcal{N}(0, \sigma_x^2)$, and $Y \sim \mathcal{N}(0, \sigma_y^2)$
- $f_{X|Y}(x|y)$ would be a Gaussian density with mean $y\rho\frac{\sigma_x}{\sigma_y}$ and variance $\sigma_x^2(1-\rho^2)$.
- Exercise for you – show all this starting with the joint density we have

- Let $\mathbf{X} = (X_1, \cdots, X_n)^T$ be jointly Gaussian.
- Then we call $\mathbf{X}$ as a Gaussian vector.
- It is possible that $X_i, i = 1, \cdots, n$ are individually Gaussian but $\mathbf{X}$ is not a Gaussian vector.
- For example, X, Y may be individually Gaussian but their joint density is not the bivariate normal density.
- Gaussian vectors have some special properties. (E.g., uncorrelated implies independence)
- Important to note that 'individually Gaussian' does not mean 'jointly Gaussian'

- The multi-dimensional Gaussian density has some important properties.
- We have seen some of them earlier.
- If $X_1, \cdots, X_n$ are jointly Gaussian then they are independent if they are uncorrelated.
- Suppose $X_1, \cdots, X_n$ be jointly Gaussian and have zero means. Then there is an orthogonal transform $\mathbf{Y} = A\mathbf{X}$ such that $Y_1, \cdots, Y_n$ are jointly Gaussian and independent.
- Another important property is the following
- $X_1, \cdots, X_n$ are jointly Gaussian if and only if $\mathbf{t}^T\mathbf{X}$ is Gaussian for for all non-zero $\mathbf{t} \in \Re^n$.
- We will prove this using moment generating functions