Week 13

4.4: Mixture & Hierarchical models (cont.)

Thm. If X and Y are any RV's, then

$$\underline{EX = E[\,E(X|Y)\,]}$$

provided that the expectations exist.

("law of iterated expectation")

Proof. Suppose $(X,Y) \sim f(x,y)$ - continuous setting. Then

$$EX = \iint x\, f_{XY}(x,y)\,dx\,dy$$

$$= \iint x\, f_{X|Y}(x|y)\, f_Y(y)\,dx\,dy$$

$$= \int \left(\int x\, f_{X|Y}(x|y)\,dx\right) f_Y(y)\,dy$$

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

$$= E[\,E(X|Y)\,]. \blacksquare$$

Def: A RV X is said to have a mixture dist. if the dist. of X depends on a quantity that is also random, e.g. (Poisson-Binomial mixture)

$$\left.\begin{array}{l} X|Y \sim \text{Binomial}(Y,p) \\ Y \sim \text{Poisson}(\lambda) \end{array}\right\}$$ mixture of Binomial distributions; mixing mechanism given by Poisson

Ex: Let $Y \sim \text{Bernoulli}(p)$; $P(Y=0)=p$

$X|Y=0 \sim \text{Normal}(\mu_1, \sigma_1^2)$

$X|Y=1 \sim \text{Normal}(\mu_2, \sigma_2^2)$

Then

$$P(X \le x) = P(X \le x, Y=0) + P(X \le x, Y=1)$$

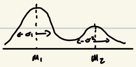
$$= P(Y=0)P(X \le x|Y=0) + P(Y=1)P(X \le x|Y=1)$$

$$= p\, F_{N(\mu_1,\sigma_1^2)}(x) + (1-p)\, F_{N(\mu_2,\sigma_2^2)}(x)$$

Differentiating w.r.t. x:

$$f_X(x) = p\, f_{N(\mu_1,\sigma_1^2)}(x) + (1-p)\, f_{N(\mu_2,\sigma_2^2)}(x)$$

The RHS is the pdf of a mixture of normal distribution:



Adding more "models" to obtain arbitrarily complex distributions:

Prop.
Suppose $Y \in \{1,\dots,k\}$, $P(Y=i)=p_i$, $i=1,\dots,k$
and $X|Y=i \sim \text{Normal}(\mu_i, \sigma_i^2)$
Then

$$f_X(x) = \sum_{i=1}^{k} p_i\, N(x|\mu_i, \sigma_i^2)$$

$$= \sum_{i=1}^{k} p_i \frac{1}{\sqrt{2\pi\sigma_i^2}} \cdot \exp\left(\frac{-1}{2\sigma_i^2}(x-\mu_i)^2\right)$$

(a mixture of k normal components)

If a prob. dist. for X may be obtained by multiple stages of cond. dist.'s, then we obtain a hierarchical model:

Ex: Let

$\lambda \to Y \to X$

$X|Y \sim \text{Binomial}(Y,p)$ (# surviving eggs)

$Y|\Lambda \sim \text{Poisson}(\Lambda)$ (# eggs laid)

$\Lambda \sim \text{Exponential}(\beta)$

The randomness of Λ captures the variation across the (insect) mothers

$$\forall x \in \mathbb{N},\ P(X=x) = \sum_y P(X=x, Y=y)$$

$$= \sum_{y=0}^{\infty} P(X=x|Y=y)P(Y=y)$$

$$\forall y \in \mathbb{N},\ P(Y=y) = P(Y=y, 0<\Lambda<\infty)$$

$$= \int_0^\infty P(Y=y|\Lambda=\lambda)\, f_\Lambda(\lambda)\, d\lambda$$

$$= \int_0^\infty e^{-\lambda} \frac{\lambda^y}{y!} \frac{1}{\beta} e^{-\lambda/\beta} \, d\lambda$$

$$= \frac{1}{\beta y!} \int_0^\infty \lambda^y e^{-\lambda(1+\frac{1}{\beta})} \, d\lambda$$

$$= \frac{1}{\beta \cancel{y!}} \Gamma(\cancel{y}+1) \left(\frac{\beta}{1+\beta}\right)^{y+1}$$

$$= \frac{1}{1+\beta} \left(\frac{\beta}{1+\beta}\right)^y$$

So $Y \sim \text{NegBinom}(p = \frac{1}{1+\beta}, r=1)$
(geometric)

Hence, the three-stage hierarchical model is equivalent mixture via (X,Y) (w/ Λ integrated out):

$$Y \sim \text{NegBinom}(p = \tfrac{1}{1+\beta}, r=1)$$
$$X|Y \sim \text{Binomial}(Y, p)$$

<u>Thm (Iterated Variance Formula)</u>: For any two RV's X, Y,

$$\boxed{\text{var} X = E[\text{var}(X|Y)] + \text{Var}[E(X|Y)]}$$

<u>P.</u> Recall $V(X) = E(X^2) - E^2(X)$

$\Rightarrow V(X|Y) = E(X^2|Y) - E^2(X|Y)$ (0)

where $E(E(X|Y)) = E(X)$

$\Rightarrow E(\text{var}(X|Y)) = EX^2 - E(E(X|Y))^2$ (1)

$\text{var}(E(X|Y)) = E(E(X|Y))^2 - (\underbrace{E\,E(X|Y)}_{(EX)^2})^2$ (2)

Thus
RHS = (1) + (2) = (0)

Ex (<u>Beta-Binomial Identity</u>): Let $X|\theta \sim \text{Binomial}(n, \theta)$, $\theta \in (0,1)$
$\theta \sim \text{Beta}(\alpha, \beta)$

Then $P(X=x) = \int P(X=x|\theta) \, f_\pi(\theta) \, d\theta$

$$= \int \binom{n}{x} \theta^x (1-\theta)^{n-x} f_\pi(\theta) \, d\theta$$

$$= \int \binom{n}{x} \theta^x (1-\theta)^{n-x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} \, d\theta$$

$$= \binom{n}{x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int \theta^{x+\alpha-1} (1-\theta)^{n+\beta-x-1} \, dx$$

$$= \frac{n!}{x!(n-x)!} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(x+\alpha)\Gamma(n-x+\beta)}{\Gamma(n+\alpha+\beta)}$$

$$= \frac{n!}{x!(n-x)!} \frac{\overbrace{(x+\alpha-1)\cdots\alpha}^{x-1} \times \overbrace{(n-x+\beta-1)\cdots\beta}^{n-2 \text{ terms}}}{\underbrace{(n+\alpha+\beta-1)\cdots(\alpha+\beta)}_{n-1 \text{ terms}}}$$

Direct calculation of EX and VX possible but quite complicated

Now $EX = E(E(X|\theta))$
$= E(n\theta)$
$= nE\theta$
$= \frac{n\alpha}{(\alpha+\beta)}$

4.5: covariance and correlation

Let (X,Y) be a bivariate vector $((X,Y) \in \mathbb{R}^2)$

e.g. $(X,Y) = (\text{height}, \text{weight})$

want to make statements like

"If $X \uparrow$ then Y tends $\uparrow$ (and vice versa)"

<u>Def (covariance)</u>: $\boxed{\operatorname{cov}(X,Y) := E(X - E(X))(Y - E(Y))}$

<u>Remarks</u>: • If $X = Y$, then $\operatorname{cov}(X,X) \equiv \operatorname{var}(X) = E[X - EX]^2$

• If $EX = EY = 0$ then $\operatorname{cov}(X,Y) = E(XY)$

<u>Correlation</u>: $\boxed{\operatorname{corr}(X,Y) := \dfrac{\operatorname{cov}(X,Y)}{\sqrt{\operatorname{var} X}\,\sqrt{\operatorname{var} Y}}}$

i.e. $\operatorname{corr}(X,Y)$ is covariance of standardized versions of X and Y

<u>Proof.</u> Let $X' = \dfrac{X - EX}{\sqrt{\operatorname{var} X}}$, $Y' = \dfrac{Y - EY}{\sqrt{\operatorname{var} Y}}$

then $\begin{cases} EX' = EY' = 0 \\ VX' = VY' = 1 \end{cases}$

By definition, $\operatorname{corr}(X',Y') = \operatorname{cov}(X',Y')$

$$= E(X'Y')$$

$$= E\, \frac{(X - EX)}{\sqrt{\operatorname{var} X}} \frac{(Y - EY)}{\sqrt{\operatorname{var} Y}}$$

$$= \frac{E(X - EX)(Y - EY)}{\sqrt{\operatorname{var} X}\,\sqrt{\operatorname{var} Y}}$$

$$= \operatorname{corr}(X,Y). \blacksquare$$

<u>Thm.</u> i) $\operatorname{cov}(X,Y) = E(XY) - (EX)(EY)$

ii) If $X \perp\!\!\!\perp Y$ then $\operatorname{cov}(X,Y) = \operatorname{corr}(X,Y) = 0$

<u>P(i).</u> $\operatorname{cov}(X,Y) = E[(X - EX)(Y - EY)]$

$$= E(XY - X(EX) - Y(EX) + (EX)(EY))$$

L.o.E

$$= E(XY) - (EX)(EY) - (EY)\cancel{(EX)} + (E\cancel{Y})(EY)$$

$$= E(XY) - (EX)(EY). \blacksquare$$

<u>Remarks</u>: • generalizes $\operatorname{var} X = E(X^2) - E^2(X)$

• conversely, $\operatorname{cov}(X,Y) = 0$ does <u>not</u> imply $X \perp\!\!\!\perp Y$

(unless (X,Y) is bivariate normal - see later)

<u>Thm.</u> $\forall a, b \in \mathbb{R}$,

$$\boxed{V(aX + bY) = a^2 \operatorname{var}(X) + b^2 \operatorname{var}(Y) + 2ab \operatorname{cov}(X,Y)}$$

<u>Remarks</u>: It can be shown that

$$|\operatorname{cov}(X,Y)| \le \sqrt{\operatorname{var}(X)}\,\sqrt{\operatorname{var}(Y)}$$

<u>Thm.</u> If $\operatorname{var} X \operatorname{var} Y \neq 0$, then

$$-1 \le \operatorname{corr}(X,Y) \le 1$$

<u>P.</u> Suppose $\operatorname{corr}(X,Y) = 1 \Rightarrow \operatorname{cov}(X,Y) = \sqrt{\operatorname{var} X}\,\sqrt{\operatorname{var} Y}$. Then

$$V(aX + Y) = a^2 VX + 2a \operatorname{cov}(X,Y) + VY$$

$$= a^2 \operatorname{var} X + 2a \sqrt{\operatorname{var} X \operatorname{var} Y} + \operatorname{var} Y$$

$$= (a\sqrt{\operatorname{var} X} + \sqrt{\operatorname{var} Y})^2$$

For $a = -\sqrt{\operatorname{var} Y} / \sqrt{\operatorname{var} X}$, RHS $= 0$

$\Rightarrow \operatorname{var}(aX + Y) = 0 \Rightarrow aX + Y = \text{const.}$

$Y - (-a)X = \text{const.}$ "w prob. 1"

$\underline{\text{Thm}}$. Suppose $\mathrm{var}X \neq 0$, $\mathrm{var}Y \neq 0$.

i) If $\mathrm{corr}(X,Y) = 1$, then there are constants $b>0$ and c s.t.

$$Y = bX + c \text{ w/ prob. } 1$$

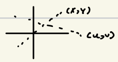
ii) If $\mathrm{corr}(X,Y) = -1$, then

$$Y = bX + c \text{ w/ prob. } 1$$

for some constants $b<0$, c.

Remarks:

if $X \perp\!\!\!\perp Y$, then $\mathrm{corr}(X,Y)=0$



$\mathrm{corr}(X,Y)=1 \Rightarrow Y = bX+c, b>0$

$\mathrm{corr}(X,Y)=-1 \Rightarrow Y = bX+c, b<0$

Ex: if $\mathrm{corr}(X,Y) \approx 1 \Leftrightarrow Y \approx bX+c, b>0$

$\mathrm{corr}(X,Y) \approx -1 \Leftrightarrow Y \approx bX+c, b<0$

Let $X \sim \mathrm{Unif}(0,1)$, $Z \perp\!\!\!\perp X$

$Z \sim \mathrm{unif}(0, 1/10)$

$Y = X + Z$

$$\begin{aligned}
\mathrm{cov}(X,Y) &= EXY - EXEY \\
&= EX(X+Z) - EX(EX+EZ) \\
&\overset{\text{L.O.E}}{=} EX^2 + EXZ - (EX)^2 - (EX)(EZ) \\
&\qquad X \perp\!\!\!\perp Z \Rightarrow EXZ - (EX)(EZ) = 0 \\
&= \mathrm{var}X \\
&= 1/12
\end{aligned}$$

Hence, $\mathrm{corr}(X,Y) = \dfrac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}X\,\mathrm{var}Y}} = \dfrac{1/12}{\sqrt{1/12}\sqrt{1/12 + 1/1200}}$

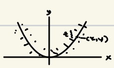
$$= \frac{1}{\sqrt{1+\frac{1}{100}}} = \left(\frac{100}{101}\right)^{1/2} \approx 1$$

Remark: $\mathrm{corr}(X,Y) = 0$ does not imply $X \perp\!\!\!\perp Y$.

Ex: Let $X \sim \mathrm{unif}(-1,1)$

$Z \sim \mathrm{unif}(0, 1/10)$, $Z \perp\!\!\!\perp X$

$Y = X^2 + Z$



$$\begin{aligned}
\text{then } \mathrm{cov}(X,Y) &= EXY - (EX)(EY) \\
&= EX(X^2+Z) - (EX)(EX^2 + EZ) \\
&= EX^3 + EXEZ - (EX)(EX^2) - (EX)(EZ) \\
&= EX^3 - EXEX^2 \\
&\qquad \text{where } X \text{ symmetric around } 0 \\
&= 0 - 0 = 0
\end{aligned}$$

Bivariate Normal Distribution

Def: $Z = (X,Y) \sim N(\mu, \Sigma)$ $\Big|\begin{array}{l}\mu \in \mathbb{R}^2 \\ \Sigma \text{ positive definite, symmetric } \in \mathbb{R}^{2\times 2}\end{array}$

$$\boxed{f_{XY}(x,y) := f_Z(z) := \frac{1}{\left((2\pi)^2 |\Sigma|\right)^{1/2}} \exp\left\{ -\frac{1}{2}(z-\mu)^T \Sigma^{-1} (z-\mu) \right\}} \quad (*)$$

Remark: $\mu = (\mu_1, \mu_2)^T$, $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$

μ: mean vector, Σ: covariance matrix of Z : $\begin{cases} EZ = \mu \\ \mathrm{cov}Z = \Sigma \end{cases}$

$EZ = \mu$ means $E\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \Leftrightarrow \begin{cases} EX = \mu_1 \\ EY = \mu_2 \end{cases}$

$\mathrm{cov}Z = \Sigma$ means $E[Z - EZ][Z-EZ]^T = \Sigma$

$$\Leftrightarrow E\begin{bmatrix} X - EX \\ Y - EY \end{bmatrix}\begin{bmatrix} X - EX \\ Y - EY \end{bmatrix}^T = \Sigma$$

$$\Leftrightarrow E\begin{bmatrix} (X-EX)^2 & (X-EX)(Y-EY) \\ (Y-EY)(X-EX) & (Y-EY)^2 \end{bmatrix} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

$$\Leftrightarrow \begin{cases} \Sigma_{11} = \mathrm{var}X \\ \Sigma_{22} = \mathrm{var}Y \\ \Sigma_{12} = \Sigma_{21} = \mathrm{cov}(X,Y) \end{cases} \quad \text{where } \mathrm{cov}Z = E[(Z-EZ)(Z-EZ)^T]$$

a remarkable property of the normal dist.

Due to the identities:

if $f_{XY}(x,y) = f_Z(z) = \frac{1}{\sqrt{(2\pi)^2 |\Sigma|}} \exp\left\{ -\frac{1}{2} (z-\mu)^T \Sigma^{-1} (z-\mu) \right\}$

then:

$$\boxed{\begin{aligned} &\int_{\mathbb{R}}\int_{\mathbb{R}} f_{XY}(x,y)\,dx\,dy = 1 \\ &\iint x f_{XY}(x,y)\,dx\,dy = \mu_1 \\ &\iint y f_{XY}(x,y)\,dx\,dy = \mu_2 \\ &\iint (z-\mu)(z-\mu)^T f_Z(z)\,dz = \Sigma \end{aligned}}$$

remark: follows from identities for (univariate) normal dist.

Translating (*) into univariate forms (don't need to remember):

$$|\Sigma| = \begin{vmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{vmatrix} = \Sigma_{11}\Sigma_{22} - \Sigma_{21}\Sigma_{22}$$

$= \sigma_X^2 \sigma_Y^2 - \sigma_{XY}^2$ where σ_{XY}: $\mathrm{cov}(X,Y)$

$= \sigma_X^2 \sigma_Y^2 (1 - \rho_{XY}^2)$

where $\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \mathrm{corr}(X,Y)$

$\mu = (\mu_1, \mu_2)^T \equiv (\mu_X, \mu_Y)^T$

$\Sigma^{-1} = \frac{1}{|\Sigma|} \begin{pmatrix} \Sigma_{22} & -\Sigma_{21} \\ -\Sigma_{12} & \Sigma_{11} \end{pmatrix}$, $A^{-1} = \frac{1}{|A|} A$

$$= \frac{1}{\sigma_X^2 \sigma_Y^2 (1-\rho_{XY}^2)} \begin{pmatrix} \sigma_Y^2 & -\sigma_{XY} \\ -\sigma_{XY} & \sigma_X^2 \end{pmatrix}$$

$$= \frac{1}{1-\rho_{XY}^2} \begin{pmatrix} \frac{1}{\sigma_X^2} & \frac{-\sigma_{XY}}{\sigma_X^2 \sigma_Y^2} \\ \frac{-\sigma_{XY}}{\sigma_X^2 \sigma_Y^2} & \frac{1}{\sigma_Y^2} \end{pmatrix}$$

$$\frac{-1}{2}(z-\mu)^T \Sigma^{-1} (z-\mu) = -\frac{1}{2(1-\rho_{XY}^2)} \left\{ \frac{1}{\sigma_X^2}(x-\mu_X)^2 - \frac{2\sigma_{XY}}{\sigma_X^2 \sigma_Y^2}(x-\mu_X)(y-\mu_Y) + \frac{1}{\sigma_Y^2}(y-\mu_Y)^2 \right\}$$

where $2\rho_{XY}\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) = \frac{2\sigma_{XY}}{\sigma_X^2 \sigma_Y^2}(x-\mu_X)(y-\mu_Y)$

hence,

$$f_{XY}(x,y) = \frac{1}{2\pi \sigma_X \sigma_Y (1-\rho_{XY}^2)^{1/2}} \times \exp\left\{ \frac{-1}{2(1-\rho_{XY}^2)} \left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho_{XY}\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 \right] \right\}$$

Thm. If $\mathrm{cov}(X,Y) = 0$ and $(X,Y) \sim$ bivariate Normal then $X \perp\!\!\!\perp Y$

P. $\mathrm{cov}(X,Y) = 0 \Rightarrow \rho_{XY} = 0$

From (*), $f_{XY}(x,y)$ factorizes into a product of $g(x)h(y)$, so $X \perp Y$. ∎

Facts: If $Z = (X,Y)^T \sim N(\mu, \Sigma)$ $\Bigg| \begin{aligned} &\mu = (\mu_1, \mu_2)^T \\ &\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \end{aligned}$

then the marginal of X and Y are normal too:

$X \sim N(\mu_1, \Sigma_{11})$

$Y \sim N(\mu_2, \Sigma_{22})$

and $\mathrm{cov}(X,Y) = \Sigma_{12} = \Sigma_{21}$

the conditional distributions are also normal:

$Y|_{X=x} \sim N(E(Y|X=x), \mathrm{var}(Y|X=x))$

where $E(Y|X=x) = \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x-\mu_1)$

$\mathrm{var}(Y|X=x) = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$