Week 12

Recap: $(X,Y) \sim f_{XY}$ (joint pdf)

marginal pdf: $= \begin{cases} f_X(x) = \int f_{XY}(x,y)\,dy \\ f_Y(y) = \int f_{XY}(x,y)\,dx \end{cases}$

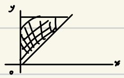
conditional pdf: $= f_{Y|X}(y|x) = \dfrac{f_{XY}(x,y)}{f_X(x)}$

$$P(Y \in B \mid X = x) = \int_B f_{Y|X}(y|x)\,dy$$

note: $P(Y \in B \mid X \in A) \neq \int_B \int_A f_{Y|X}(y|x)\,dx\,dy$?why

$f_{XY}(x,y) = f_X(x)\, f_{Y|X}(y|x)$ for x,y s.t. $f_X(x) > 0$

Ex: Let $(X,Y) \sim f_{XY}(x,y) = \begin{cases} e^{-y}, & 0<x<y<\infty \\ 0, & \text{otherwise} \end{cases}$



check $f_{XY}(x,y) \geq 0 \ \forall x,y \in \mathbb{R}^2$.

$$\int_{\mathbb{R}^2} f_{XY}(x,y)\,dx\,dy = \int_0^\infty \int_0^y e^{-y}\,dx\,dy = \int_0^\infty y e^{-y}\,dy$$

$$= \int_0^\infty (-y)\,de^{-y}$$

$$\overset{\text{IBP}}{=} \int_0^\infty e^{-y}\,dy = 1$$

$\Rightarrow$ valid joint pdf. ∎

Marginal computation: $f(x) = 0 \ \forall x \leq 0$

for $x > 0$

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy = \int_x^\infty f_{XY}(x,y)\,dy$$

$$= \int_x^\infty e^{-y}\,dy = e^{-y}\Big|_x^\infty = e^{-x}$$

for $y > 0$:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx = \int_0^y f_{XY}(x,y)\,dx = \int_0^y e^{-y}\,dx = ye^{-y}$$

$f_Y(y) = 0 \ \forall y \leq 0$

conditional pdf: If $y \geq x > 0$, then

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} = \begin{cases} \dfrac{e^{-y}}{e^{-x}} = e^{x-y} \\ 0, & y < x \end{cases}$$

If $x \in (0,y)$:

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{e^{-y}}{ye^{-y}} = \frac{1}{y} \text{ and } 0 \text{ otherwise}$$

conditional expectations

$$E(Y \mid X = x) = \int y\, f_{Y|X}(y|x)\,dy = \int_x^\infty y e^{x-y}\,dy$$

$$= e^x \int_x^\infty y e^{-y}\,dy = x+1$$

$$E(Y^2 \mid X = x) = \int_x^\infty y^2 e^{x-y}\,dy = e^x \int_x^\infty y^2 e^{-y}\,dy = x^2+2x+2$$

Hence $\operatorname{var}(Y \mid X = x) = x^2+2x+2-(x+1)^2 = 1$

Ex: Given two light bulbs w/ life lengths $X, Z \geq 0$

Let $X, Z \overset{iid}{\sim} \text{Exp}(\lambda)$

Let $Y = X + Z$: time bulb 2 burns out

$$P(Y \leq y \mid X = x) = P(Z \leq y - x \mid X = x)$$

$$= P(Z \leq y - x) \text{ since } Z \perp X$$

$$= F_Z(y-x) = 1 - e^{-\lambda(y-x)} \quad \text{if } y > x$$

Hence

$$f_{Y|X}(y|x) = \frac{d}{dy}\left(1 - e^{-\lambda(y-x)}\right) = \lambda e^{-\lambda(y-x)}$$

<u>note</u>. If $\lambda = 1$, this gives same cond. pdf as in previous example

Then,

$$f_{XY}(x,y) = f_X(x) f_{Y|X}(y|x)$$

$$= \lambda^2 e^{-\lambda y} \quad \text{if } y > x > 0$$

and $f_{XY}(x,y) = 0$ otherwise

4.2b: Independence

<u>Def</u>: X and Y are independent RV's if, $\forall x, y \in \mathbb{R}$

$$\underline{f_{XY}(x,y) = f_X(x) f_Y(y)}$$

<u>remarks</u>: 1) If $X \perp\!\!\!\perp Y$ then $f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{\cancel{f_X(x)} f_Y(y)}{\cancel{f_X(x)}} = f_Y(y)$

(does not depend on x)

Moreover, $\forall A, B \subset \mathbb{R}$,

$$P(Y \in B \mid X \in A) = \frac{\int_B \int_A f_{XY}(x,y)\,dx\,dy}{\int_A f_X(x)\,dx} = \frac{\int_B \int_A f_X(x) f_Y(y)\,dx\,dy}{\int_A f_X(x)\,dx} = \frac{\left(\int_B f_Y(y)\,dy\right)\left(\int_A f_X(x)\,dx\right)}{\int_A f_X(x)\,dx} = \int_B f_Y(y)\,dy = P(Y \in B)$$

Hence the event $\{Y \in B\}$ is independent of $\{X \in A\}$ $\forall A, B$

<u>Remark</u>: to verify independence, need to check the above identity for all x, y (or all A, B)

to show non-independence, need to identify a pair (x,y) or (A,B) where identity not satisfied.

<u>Ex</u>: Suppose $f_{XY}(x,y)$ is given by the table

x \ y	1	2	3
10	1/10	1/5	1/5
20	1/10	1/10	3/10

all numbers sum to 1

$$f_X(10) = f_{XY}(10,1) + f_{XY}(10,2) + f_{XY}(10,3)$$

$$= 1/10 + 1/5 + 1/5 = 1/2$$

$$f_Y(3) = 1/5 + 3/10 = 1/2$$

But $f_{XY}(10,3) = 1/5 \neq f_X(10) \times f_Y(3)$

$\Rightarrow X, Y$ not independent.

> **Lemma**: Let $(X,Y) \sim f_{XY}$. Then $X \perp\!\!\!\perp Y$ i.f.f. $\exists$ functions $g(x)$ and $h(y)$ s.t.
>
> $$f_{XY}(x,y) = g(x)h(y) \quad \forall x, y \in \mathbb{R}$$

<u>Proof</u>: "only if": when $X \perp\!\!\!\perp Y$, $f_{XY}(x,y) = f_X(x) f_Y(y)$
$= g(x) \cdot h(y)$

"if": Given that $f_{XY}(x,y) = g(x)h(y)$,

$$1 = \iint_{\mathbb{R}\,\mathbb{R}} f_{XY}(x,y)\,dx\,dy = \iint g(x)h(y)\,dx\,dy$$

$$= \underbrace{\left(\int g(x)\,dx\right)}_{c} \underbrace{\left(\int h(y)\,dy\right)}_{d} \Rightarrow cd = 1$$

Now

$$f_X(x) = \int_{\mathbb{R}} f_{XY}(x,y)\,dy = g(x)\,d$$

$$f_Y(y) = \int_{\mathbb{R}} f_{XY}(x,y)\,dx = h(y)\,c$$

Hence

$$f_X(x) f_Y(y) = g(x)h(y)cd = g(x)h(y) = f_{XY}(x,y). \ \blacksquare$$

<u>Ex</u>: If $f_{XY}(x,y) = \frac{1}{384} x^2 y^4 e^{-y-(x/2)}$, $x, y > 0$

$$= \frac{1}{384} x^2 e^{-x/2}\, y^4 e^{-y}$$

$\Rightarrow X \perp\!\!\!\perp Y$.

Thm. If $X \perp\!\!\!\perp Y$, then

1) $\forall A \subset \mathbb{R}, B \subset \mathbb{R}$
$P(X \in A, Y \in B) = P(X \in A) P(Y \in B)$

2) $\forall$ function $g(x)$ - only of x
$h(y)$ - only of y
$E[g(X)h(Y)] = Eg(X)\, Eh(Y)$

Thm. If $X \perp\!\!\!\perp Y$, w/ mgf M_X and M_Y, then RV $Z := X+Y$ has the mgf
$M_Z(t) = M_X(t) \cdot M_Y(t) \quad \forall t$

P. $M_Z(t) = Ee^{tZ} = Ee^{t(X+Y)}$
$= Ee^{tX} e^{tY}$
$= Ee^{tX} Ee^{tY}$, $X \perp\!\!\!\perp Y$
$= M_X(t) M_Y(t)$. ∎

Thm. The sum of two independent normal RV's is again Normal

P. Let $X \sim \text{Normal}(0,1)$. Then
$$M_X(t) = Ee^{tX} = \int_{\mathbb{R}} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$$
$$= \frac{1}{\sqrt{2\pi}} \int e^{tx - \frac{1}{2}x^2} dx$$, completing the square
$$= \frac{1}{\sqrt{2\pi}} \int e^{-\frac{1}{2}(x-t)^2} e^{\frac{1}{2}t^2} dx$$
$$= e^{\frac{1}{2}t^2}$$

If $X \sim \text{Normal}(\mu, \sigma^2)$, then
$Z := \frac{1}{\sigma}(X-\mu) \sim N(0,1)$, so
$$M_X(t) = Ee^{tX}$$
$$= Ee^{t(\mu + \sigma Z)}$$
$$= e^{t\mu} Ee^{t\sigma Z}$$
$$= e^{t\mu} M_Z(t\sigma)$$
$$= e^{t\mu + \frac{1}{2}t^2\sigma^2}$$

Now suppose $X \sim \text{Normal}(\mu_1, \sigma_1^2)$
$Y \sim \text{Normal}(\mu_2, \sigma_2^2)$
$X \perp\!\!\!\perp Y$.

Let $Z = X+Y$. Then,
$$M_Z(t) = M_X(t) M_Y(t)$$, since $X \perp\!\!\!\perp Y$
$$= e^{t\mu_1 + \frac{1}{2}t^2\sigma_1^2} e^{t\mu_2 + \frac{1}{2}t^2\sigma_2^2}$$
$$= e^{t(\mu_1+\mu_2) + \frac{1}{2}t^2(\sigma_1^2 + \sigma_2^2)}$$

i.e. the mgf of $N(\mu_1+\mu_2, \sigma_1^2 + \sigma_2^2)$

Now, since a RV's dist. is determined by its mgf,

$Z \sim \text{Normal}(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$. ∎

4.3: Bivariate Transformation

Let (X,Y) be a bivariate random vector
$g: \mathbb{R}^2 \to \mathbb{R}^2$
$g(x,y) := (g_1(x,y), g_2(x,y)) \in \mathbb{R}^2$

Then $(U,V) := g(X,Y)$ is a bivariate random vector
$\forall A \in \mathbb{R}^2$, $P((U,V) \in A) = P((X,Y) \in g^{-1}(A))$
where $g^{-1}(A) = \{(x,y) \mid g(x,y) \in A\}$

Discrete case: If (X,Y) is discrete, then so is (U,V)
$$f_{UV}(u,v) = \sum_{\substack{x,y: \\ g_1(x,y)=u \\ g_2(x,y)=v}} f_{XY}(x,y)$$

Ex: Let $X \sim \text{Poisson}(\theta)$, $Y \sim \text{Poisson}(\lambda)$, $X \perp\!\!\!\perp Y$.

First Approach: $M_X(t) = Ee^{tX} = \sum_{x=0}^{\infty} e^{tx} e^{-\theta} \frac{\lambda^x}{x!}$
$$= \sum_{x=0}^{\infty} e^{-\theta} \frac{(e^t)^x}{x!}$$
$$= e^{\theta(e^t-1)}.$$

Hence $M_{X+Y}(t) = M_X(t) M_Y(t)$ since $X \perp\!\!\!\perp Y$
$$= e^{\lambda(e^t-1)} e^{\theta(e^t-1)}$$
$$= e^{(\lambda+\theta)(e^t-1)}$$
which is the mgf for Poisson$(\lambda+\theta)$.

Second Approach: Write $\binom{U}{V} = g\left(\binom{X}{Y}\right) := \binom{X+Y}{Y}$

Thus, for $u \ge v$,

$$f_{UV}(u,v) = \sum_{\substack{x,y \\ y=v \\ x+y=u}} f_{XY}(x,y)$$

$$= f_{XY}(u-v, v)$$

$$= f_X(u-v) f_Y(v) \quad \text{since } X \perp\!\!\!\perp Y$$

$$= e^{-\theta} \frac{\theta^{u-v}}{(u-v)!} \, e^{-\lambda} \frac{\lambda^v}{v!}$$

$$= e^{-(\theta+\lambda)} \frac{\theta^{u-v}}{(u-v)!} \cdot \frac{\lambda^v}{v!}$$

Obtain $f_U(u)$ by marginalization:

$$f_U(u) = \sum_v f_{UV}(u,v) \quad \text{since } f_{UV}(u,v) = 0 \text{ if } v > u$$

$$= e^{-(\theta+\lambda)} \sum_{v=0}^{u} \theta^{u-v} \lambda^v \frac{1}{(u-v)!\,v!}$$

$$= e^{-(\theta+\lambda)} \frac{1}{u!} \sum_{v=0}^{u} \theta^{u-v} \lambda^v \binom{u}{v}$$

$$= e^{-(\theta+\lambda)} \frac{1}{u!} (\theta+\lambda)^u \quad \text{via binomial formula}$$

Hence, $U = X+Y \sim \text{Poisson}(\theta+\lambda)$.

<u>Continuous case</u>: change-of-var formula.

Let (X,Y) - continuous bivariate vector
$(X,Y) \sim f_{XY}$

Let $g: \mathbb{R}^2 \to \mathbb{R}^2$ be a one-to-one mapping
i.e. $g^{-1}(\{u,v\})$ has at most one element

Define $(U,V) = g(X,Y)$; what is $f_{UV}(u,v)$?

Let $A = (\{(x,y)\}, f_{XY}(x,y) > 0)$, support of f_{XY}

$B = g(A) := \{g(x,y) \mid (x,y) \in A\}$

write $\binom{u}{v} := g(x,y) := \begin{pmatrix} g_1(x,y) \\ g_2(x,y) \end{pmatrix}$

which has the inverse function
$\binom{x}{y} = h(u,v) := g^{-1}(u,v) := \begin{pmatrix} h_1(u,v) \\ h_2(u,v) \end{pmatrix}$

Then the pdf $f_{UV}(u,v)$ is given by the

$$\boxed{\underline{\text{change-of-var formula}} : f_{UV}(u,v) = f_{XY}(h_1(u,v), h_2(u,v)) \, |J|}$$

where J: determinant of the Jacobian matrix:

$$J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} \frac{\partial h_1}{\partial u} & \frac{\partial h_1}{\partial v} \\ \frac{\partial h_2}{\partial u} & \frac{\partial h_2}{\partial v} \end{vmatrix} = \frac{\partial h_1}{\partial u}\frac{\partial h_2}{\partial v} - \frac{\partial h_1}{\partial v}\frac{\partial h_2}{\partial u}$$

<u>EX</u>: Let $X \sim N(0,1)$, $Y \sim N(0,1)$, $X \perp\!\!\!\perp Y$

Via mgf, we showed $\begin{cases} X+Y \sim N(0,2) \\ X-Y \sim N(0,2) \end{cases}$

Now, via change-of-var formula,

let $\begin{cases} U = X+Y \\ V = X-Y \end{cases}$

the mapping $\begin{cases} u = g_1(x,y) := x+y \\ v = g_2(x,y) := x-y \end{cases}$ is one-to-one and admits the inverse:

$$\begin{cases} x = h_1(u,v) = \frac{1}{2}(u+v) \\ y = h_2(u,v) = \frac{1}{2}(u-v) \end{cases}$$

Hence, $J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{vmatrix} = -\frac{1}{4} - \frac{1}{4} = -\frac{1}{2}$

Thus, $f_{UV}(u,v) = f_{XY}(\frac{1}{2}(u+v), \frac{1}{2}(u-v)) \cdot \frac{1}{2}$

$$= f_X(\tfrac{1}{2}(u+v)) f_Y(\tfrac{1}{2}(u-v)) \cdot \tfrac{1}{2}, \quad X \perp\!\!\!\perp Y$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{1}{2}(u+v))^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{1}{2}(u-v))^2} \cdot \frac{1}{2}$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-\frac{1}{4}u^2} \cdot \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-\frac{1}{4}v^2}$$

$f_{UV}(u,v)$ factorizes $\Rightarrow U \perp\!\!\!\perp V$, i.e. $(X+Y) \perp\!\!\!\perp (X-Y)$ and

$U \sim N(0,2)$
$V \sim N(0,2)$

When the transformation $(X,Y) \to g(X,Y) \in \mathbb{R}^2$ is many-to-one,

we may partition the support of f_{XY}

$$A = \{(x,y) \mid f_{XY}(x,y) > 0\}$$

into disjoint subsets

$$A = A_1 \cup \cdots \cup A_k$$

s.t. g is one-to-one from each A_i to $g(A_i)$, $i=1,\dots,k$

Prop. Let $h_i = (h_{i1}, h_{i2})$ be the inverse function of g restricted to domains $A_i \to g(A_i)$ and J_i the corresponding Jacobian. Then

$$\underline{f_{UV}(u,v) = \sum_{i=1}^{k} f_{XY}(h_{i1}(u,v), h_{i2}(u,v))\,|J_i|}$$

Ex: Let $X, Y \sim N(0,1)$, $X \perp\!\!\!\perp Y$.

Suppose $\begin{cases} U = X/Y \\ V = |Y| \end{cases}$

note (x,y) and $(-x,-y)$ map to the same (u,v)

so partition $A = \mathbb{R}^2 = \underbrace{\mathbb{R}\times\mathbb{R}_+}_{A_1} \cup \underbrace{\mathbb{R}\times\mathbb{R}_-}_{A_2} \cup \underbrace{\mathbb{R}\times\{0\}}_{A_3}$

If $(x,y) \in A_3$ $f_{XY}(x,y) = f_X(x)f_Y(y)$ but (u,v) is not defined; thus, not of interest since $P((X,Y)\in A_3) = 0$

For A_1: $\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} x/y \\ y \end{pmatrix} \Rightarrow \begin{cases} x = uv \\ y = v \end{cases} \Rightarrow \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} uv \\ v \end{pmatrix}$

$$J_1 = \begin{vmatrix} v & u \\ 0 & 1 \end{vmatrix} = v$$

For A_2: $\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -uv \\ -v \end{pmatrix} \Rightarrow J_2 = \begin{vmatrix} -v & -u \\ 0 & -1 \end{vmatrix} = v$

Hence, $f_{UV}(u,v) = f_{XY}(h_1(u,v))|J_1| + f_{XY}(h_2(u,v))|J_2|$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(uv)^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}v^2} |v| + \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(-uv)^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(-v)^2} |v|$$

$$= \underline{\frac{1}{\pi} e^{-\frac{1}{2}v^2(u^2+1)} v \quad , \text{for } v > 0}$$

$f_{UV}(u,v) = 0 \ \forall\ v < 0$. Thus,

$$f_U(u) = \int_0^\infty \frac{1}{\pi} e^{-\frac{1}{2}v^2(u^2+1)} v \, dv$$

$$= \frac{1}{\pi(u^2+1)} e^{-\frac{1}{2}v^2(u^2+1)} \Big|_\infty^0$$

$$= \frac{1}{\pi(u^2+1)} \quad \text{for all } u \in \mathbb{R}$$

$\Rightarrow \underline{U = X/Y \sim \text{Cauchy}(1)}$

4.4: Mixture & Hierarchical Models

<u>Idea</u>: compute families of dist.'s in terms of simpler models of dist.'s, which are constructed in a hierarchical way

<u>Ex</u>: Insect lays large number of eggs
Each insect survives w/ prob. p
Let X: # eggs themselves

<u>Q</u>: How to model behavior of X

Bring in (latent) RV Y which represents the number of eggs laid;

assume $Y \sim \text{Poisson}(\lambda)$
$X|Y \sim \text{Binomial}(Y,p)$
↳ mixture of binomial dist.

This defines a proper joint dist. for (X,Y) which induces a proper (marginal) dist. for X: $f_{XY}(x,y) = f_Y(y) f_{X|Y}(x|y)$

i.e. $\forall\ x,y \in \mathbb{N}$, $P(X=x, Y=y) = P(Y=y)P(X|Y)$

So,

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

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

$$= \sum_{y=0}^{\infty} \frac{e^{-\lambda}\lambda^y}{y!}\binom{y}{x} p^x(1-p)^{y-x}, \quad y \geq x$$

$$= \sum_{y=x}^{\infty} \frac{e^{-\lambda}\lambda^y}{y!} \frac{y!}{(y-x)!\,x!} p^x(1-p)^{y-x}$$

$$= \frac{\lambda^x p^x}{x!} e^{-\lambda} \sum_{y=x}^{\infty} \frac{\lambda^{y-x}}{(y-x)!}(1-p)^{y-x}$$

letting $t = y - x$

$$= \frac{\lambda^x p^x}{x!} e^{-\lambda} \sum_{t=0}^{\infty} \frac{\lambda^t (1-p)^t}{t!}$$

$$= \frac{\lambda^x p^x}{x!} e^{-\lambda} e^{\lambda(1-p)}$$

$$= \frac{(\lambda p)^x}{x!} e^{-\lambda p} \sim \text{Poisson}(\lambda p)$$

$$\Rightarrow \begin{cases} EX = \lambda p \\ E[X \mid Y=y] = yp \quad \forall y \in \mathbb{N} \end{cases}$$

Note: conditional expectation $E(X|Y)$ is a RV and hence

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

$$\Rightarrow E(E(X|Y)) = E(Yp) = pE(Y) = \lambda p$$

Remarks:
- Usually, marginal dist. for X may not share the same family as that of the latent Y
- latent variable Y helps to explain the role of parameters λ and p