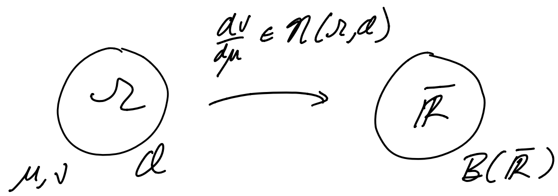
Lecture 19:
Radon-Nikodym Derivatives and Lebesgue decomposition

Recall that if μ, ν are two measures on $(\Omega, \mathcal{A})$ then $\frac{d\nu}{d\mu}$ was notation for the density of ν w.r.t μ when such a thing exists.



such that $\nu(A) = \int_A \frac{d\nu}{d\mu}\, d\mu \quad \forall A \in \mathcal{A}$.

We never had a general thm to show when $\frac{d\nu}{d\mu}$ exists. This will come from the Radon-Nikodym Thm.

This theorem is also related to the existance of conditional expected value. Here is the heuristic:

Let X and Y be two r.v.s on $(\Omega, \mathcal{A}, P)$. Suppose $X \in \mathcal{M}(\Omega, \mathcal{A})$.

In undergrad we learned

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

Indeed for any $A \in \mathcal{A}$ we have

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

So that

$$\int_A X\,dP = \int_\Omega E(I_A X|Y)\,dP \quad \forall A \in \mathcal{A}$$

Also notice that the result "characterizing ⓜ funs of ⓜ funs" from lecture 9 implies

$$A \in \sigma\langle Y\rangle \iff I_A \text{ ⓜ } \sigma\langle Y\rangle$$
$$\iff I_A(\omega) = g(Y(\omega)) \text{ for some ⓜ } g$$

$\therefore$ If $A \in \sigma\langle Y\rangle$ then I_A can be pulled out of $E(I_A X|Y)$ and we have

$$\int_A X\,dP = \int_A E(X|Y)\,dP \quad \forall A \in \sigma\langle Y\rangle$$

In otherwords $E(X|Y)$ appears to be the density of the measure $\int_\cdot X\,dP$ on $(\Omega, \sigma\langle Y\rangle)$ w.r.t. $P|_{\sigma\langle Y\rangle}$

$$\frac{d \int_\cdot X\,dP|_{\sigma\langle Y\rangle}}{d\, P|_{\sigma\langle Y\rangle}} = E(X|Y).$$

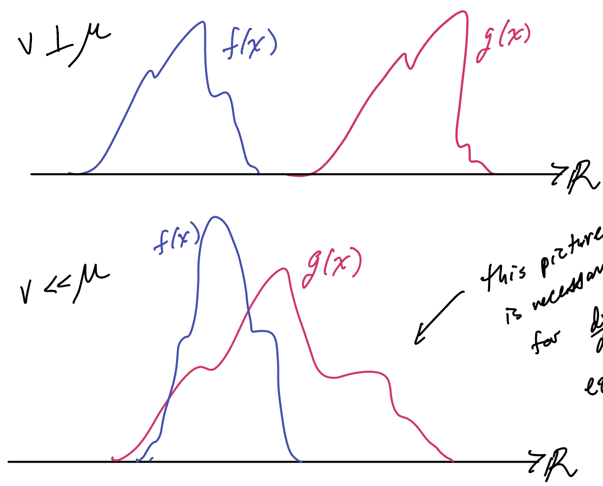
Definition: If ν & μ are measures on a measurable space $(\Omega, \mathcal{A})$ then

(i) $\nu \perp \mu$ iff $\exists A \in \mathcal{A}$ s.t.
$$\nu(A^c) = 0 = \mu(A)$$

(ii) $\nu \ll \mu$ iff $\forall A \in \mathcal{A}$
$$\mu(A) = 0 \Longrightarrow \nu(A) = 0.$$

Here is the pictures when

$$d\nu(x) = f(x)dx \quad \& \quad d\mu(x) = g(x)dx$$



this picture is necessary for $\frac{d\nu}{d\mu}$ to exist.

Before we prove the Radon-Nikodym result lets recall the following result from Lecture 11:

"Probabilits' world view"

If μ is a non-trivial & σ-finite measure on $(\Omega, \mathcal{A})$ then $\exists$ a prob measure P on $(\Omega, \mathcal{A})$ s.t. $\frac{d\mu}{dP}$ exists with the addition property that $\frac{d\mu}{dP}$ takes values in $(0, \infty)$.

Theorem: (Radon-Nikodym)

If μ & ν are two measures on $(\Omega, \mathcal{A})$ s.t. $\nu \ll \mu$ and both are σ-finite then

$$\frac{d\nu}{d\mu} \in \mathcal{M}(\Omega, \mathcal{A}) \text{ exists and is } \mu\text{-unique.}$$

Proof: If μ or $\nu \equiv 0$ the theorem is true so suppose both are non-trivial.

Since μ & σ are σ-finite the "probabilits world" implies $\exists$ probs P, Q on $(\Omega, \mathcal{A})$ s.t.

$$\frac{d\nu}{dQ} \;\&\; \frac{d\mu}{dP} \text{ exist } \& \underbrace{\text{take values in } (0,\infty)}.$$

which means exercise 3 in HWK 7 (from 235A) applies & gives

$$\therefore \quad \frac{dQ}{d\nu} = \frac{1}{d\nu/dQ} \quad \& \quad \frac{dP}{d\mu} = \frac{1}{d\mu/dP}$$

Now if $\frac{dQ}{dP}$ exists then we have

$$\frac{d\nu}{d\mu} = \frac{d\nu}{dQ}\frac{dQ}{dP}\frac{dP}{d\mu} \quad \mu\text{-a.e.}$$

by the "chain rule Thm" of Lecture 11.

Therefor all we need to do is show $\frac{dQ}{dP}$ exists.

The main idea is to define

$$W = \frac{P+Q}{2}$$

and use Riesz to get $\frac{dQ}{dW}$ & $\frac{dP}{dW}$.

Then show

$$\frac{dQ}{dP} = \frac{dQ}{dW} \Big/ \frac{dP}{dW}.$$

5

(show dQ/dP exists):

For all $X \in L_2(\Omega, \mathcal{Q}, W)$ define the following continuous linear functionals:

$$f_P(X) := \int_\Omega X\,dP = E_P(X) \overset{\text{Riesz}}{=} \langle Y_P, X\rangle_{L_2(W)}$$

$$f_Q(X) := \int_\Omega X\,dQ = E_Q(X) \overset{\text{Riesz}}{=} \langle Y_Q, X\rangle_{L_2(W)}$$

For some $Y_P, Y_Q \in L_2(\Omega, \mathcal{Q}, W)$.

To see why f_P & f_Q are continuous linear functionals over $L_2(\Omega, \mathcal{Q}, W)$ notice first that

$$\int_\Omega |X|^2 dP,\ \int_\Omega |X|^2 dQ \le 2\int_\Omega |X|^2 d\underbrace{\left(\tfrac{P+Q}{2}\right)}_{W} \quad (*)$$

So $X \in L_2(W) \Rightarrow X \in L_2(P) \cap L_2(Q)$

$\therefore$ f_P & f_Q are defined over $L_2(W)$ & clearly linear.

For continuity notice that

$$X_n \overset{L_2(W)}{\longrightarrow} X \implies \begin{matrix} X_n \overset{L_2(P)}{\longrightarrow} X \\ X_n \overset{L_2(Q)}{\longrightarrow} X \end{matrix} \quad \text{by } (*)$$

$$\implies \begin{matrix} E_P(X_n) \to E(X) \\ E_Q(X_n) \to E(X) \end{matrix} \quad \text{by } L_p \text{ convergence Thm}$$

$$\implies \begin{matrix} f_P(X_n) \to f(X) \\ f_Q(X_n) \to f(X) \end{matrix}$$

$\therefore$ Indeed, f_P & f_Q are continuous linear functionals over $L_2(W)$.

Now plug in I_A for X ($A \in \mathcal{Q}$) to get

$$f_P(I_A) = P(A) = \langle Y_P, I_A\rangle_{L_2(W)} = \int_A Y_P\,dW$$

$$f_Q(I_A) = Q(A) = \langle Y_Q, I_A\rangle_{L_2(W)} = \int_A Y_Q\,dW$$

$$\therefore\quad Y_P = \frac{dP}{dW} \quad \& \quad Y_Q = \frac{dQ}{dW}$$

modified on W-null sets so they are in $\mathcal{M}(\Omega, \mathcal{Q})$. Possible since $\int 0\,dW \le \int Y_P\,dW \Leftrightarrow 0 \le Y_P$ W-a.e. (by Thm in Lecture 11 which requires 0 or $Y_P \in L_1$ or W σ-finite)

6

Now define

$$\frac{dQ}{dP} := \frac{dQ/dW}{dP/dW} I_{\{dP/dW \neq 0\}}$$

and simply check that it serves as the density of Q w.r.t P.

Indeed, let $A \in \mathcal{Q}$ and notice

$$\underbrace{\int_A \frac{dQ}{dP}\,dP}_{\text{defined since } \frac{dQ}{dP} \in \mathcal{M}(\Omega, \mathcal{Q})} = \int_A \frac{dQ/dW}{dP/dW} I_{\{\frac{dP}{dW} \neq 0\}}\,dP$$

$$\overset{\text{"slap"}}{=} \int_A \frac{dQ/dW}{\cancel{dP/dW}} I_{\{\frac{dP}{dW} \neq 0\}} \cancel{dP/dW}\,dW$$

$$= \int_{A \cap \{\frac{dP}{dW} \neq 0\}} dQ/dW\,dW$$

$$\overset{\text{"slap"}}{=} Q\left(A \cap \left\{\tfrac{dP}{dW} \neq 0\right\}\right)$$

$$= Q(A)$$

since $P\left(A \cap \{\frac{dP}{dW} = 0\}\right) \le P\left(\frac{dP}{dW} = 0\right) = \int_{\{\frac{dP}{dW} = 0\}} \frac{dP}{dW}\,dW = 0$

and $P \gg \mu$ since $\frac{dP}{d\mu}$ exists

$\gg \nu$ by assumption

$\gg Q$ since $\frac{dQ}{d\nu}$ exists

implies $Q\left(A \cap \{\frac{dP}{dW} = 0\}\right) = 0$

QED

To recap the proof we showed

$$\frac{d\nu}{d\mu} = \frac{d\nu}{dQ}\frac{dQ}{dP}\frac{dP}{d\mu}$$

where P & Q are from "probabilists world view" which requires μ & ν σ-finite

$$= \frac{d\nu}{dQ}\frac{dQ/dW}{dP/dW}\frac{dP}{d\mu} \quad \text{for } W = \frac{P+Q}{2}$$

found by Riesz on $L_2(W)$ for $f_P(X) = E_P(X)$ & $f_Q(X) = E_Q(X)$

The following example suggests we can possibly extend the Radon-Nikodym result to the assumption μ is σ-finite rather than both μ & ν are σ-finite.

Example:

$\Omega = \mathbb{R}$

$\mathcal{A} = \mathcal{B}(\mathbb{R})$

$\mu = \mathcal{L}' :=$ Lebesgue measure

$$\nu = \infty\mu = \infty \cdot \mathcal{L}' = \begin{cases} 0 & \text{when } \mathcal{L}'(A) = 0 \\ \infty & \text{o.w.} \end{cases}$$

$\therefore$ $\nu \ll \mu$ and μ is σ-finite

but ν is not σ-finite.

Yet $\nu(A) = \int_A \infty \, d\mu$ so $\frac{d\nu}{d\mu}$ exists.

Theorem: (improved Radon-Nikodym)

If μ & ν are two measures on $(\Omega, \mathcal{A})$ s.t. $\nu \ll \mu$ and μ is σ-finite then

$\frac{d\nu}{d\mu} \in \mathcal{N}(\Omega, \mathcal{A})$ exists and is μ-unique.

Proof:

The problem here is we cannot use the "Probabilist world view" to get the existance of $\mathcal{Q}$. The plan is to find $\frac{d\nu}{dP}$ s.t.

$$\frac{d\nu}{d\mu} = \frac{d\nu}{dP} \frac{dP}{d\mu}$$

where the existance of $\frac{d\nu}{dP}$ will come from the fact that P is a finite measure.

($d\nu/dP$ exists):

Note that for any $F \in \mathcal{A}$ we can write $\nu(\cdot) = \nu(\cdot \cap F) + \nu(\cdot \cap F^c)$

we will want to find F s.t.

(i) $\nu(\cdot \cap F)$ is σ-finite

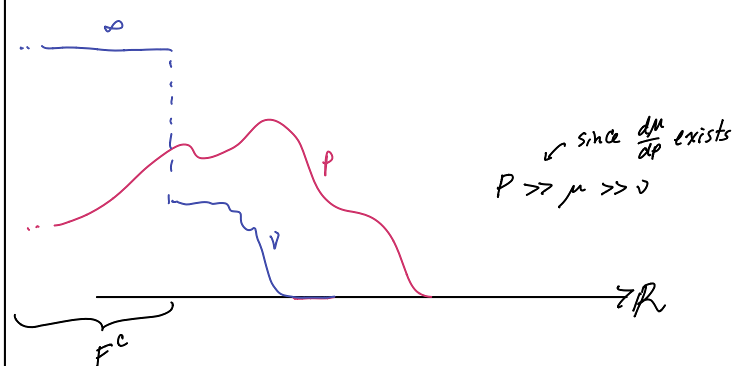
so old RD Thm applies to give $\frac{d\,\nu(\cdot \cap F)}{dP}$

(ii) $\nu(\cdot \cap F^c) = \infty P(\cdot \cap F^c)$

where this "bad piece" is covered by the last example so that $\frac{d\,\nu(\cdot \cap F^c)}{dP} := \infty I_{F^c}$

since $\int_A \infty I_{F^c} \, dP = \infty P(A \cap F^c) = \nu(A \cap F^c)$.

Here is the picture



Lets find F as the "P-biggest set s.t. ν is σ-finite over F."

Set

$$\mathcal{F} := \left\{ \bigcup_{k=1}^{\infty} A_k : \nu(A_k) < \infty,\ A_k \in \mathcal{A},\ \forall k \right\}$$

and notice that $\mathcal{F}$ is closed under countable union.

Let $m = \sup\{P(F) : F \in \mathcal{F}\}$ and choose $F = \bigcup_{k=1}^{\infty} A_k \in \mathcal{F}$ to attain the above sup.

The existance of such an F holds since

$F_n \in \mathcal{F}$ s.t. $P(F_n) \to m$ implies

$$m \xleftarrow{n\to\infty} P(F_n) \le P\Big(\underbrace{\bigcup_{n=1}^{\infty} F_n}_{\in \mathcal{F}}\Big) \le m$$

$\therefore$ sup is attained at $\bigcup_{n=1}^{\infty} F_n$.

Now we just check that

(i) $\nu(\cdot \cap F)$ is σ-finite

(ii) $\nu(\cdot \cap F^c) = \infty P(\cdot \cap F^c)$.

For (i) notice that since $F \in \mathcal{F}$ we have

$F = \bigcup_{k=1}^{\infty} A_k$ for $\nu(A_k) < \infty$ $\forall k$ and

$\therefore$ $\nu(F^c \cap F), \nu(A_1 \cap F), \nu(A_2 \cap F), \dots$

are all finite & $F^c \cup A_1 \cup A_2 \cup \dots = \Omega$

$\therefore$ $\nu(\cdot \cap F)$ is σ-finite

For (ii) notice that $\forall A \in \mathcal{A}$

$P(A \cap F^c) = 0 \implies \nu(A \cap F^c) = 0$

by $P \gg \mu \gg \nu$. Also

$P(A \cap F^c) > 0 \implies \nu(A \cap F^c) = \infty$

For suppose not.

$\therefore$ $\exists A \in \mathcal{A}$ s.t. $P(A \cap F^c) > 0$ (a)
$\nu(A \cap F^c) < \infty$ (b)

$\therefore$ $A \cap F^c \in \mathcal{F}$ by (b)

$\therefore$ $F \cup (A \cap F^c) \in \mathcal{F}$, since $F \in \mathcal{F}$ & $\mathcal{F}$ is closed under countable union

$\therefore$ $m = P(F) \overset{(a)}{<} P(F) + P(A \cap F^c)$
$= P(F \cup (A \cap F^c)) \le m$

$\therefore$ contradiction

QED

Remark: the strict inequality $\overset{(a)}{<}$ above is where we needed that P be a finite measure.

Properties of Radon-Nikodym derivatives

In this section it will be convinent to use the following (totally not standard) notation

$\nu \lll \mu$ means $\nu \ll \mu$ & μ is σ-finite

Theorem: (RND props)

Let $\nu, \mu, \sigma, \nu_1, \nu_2, \dots$ be measures on a measurable space $(\Omega, \mathcal{A})$.

(1) If $\nu_1, \nu_2 \lll \mu$ & $c_1, c_2 \ge 0$ then

$(c_1\nu_1 + c_2\nu_2) \lll \mu$ and

$$\frac{d(c_1\nu_1 + c_2\nu_2)}{d\mu} = c_1 \frac{d\nu_1}{d\mu} + c_2 \frac{d\nu_2}{d\mu} \quad \mu\text{-a.e.}$$

(2) If $\nu_1, \nu_2 \lll \mu$ then

$\nu_1 \le \nu_2$ on $\mathcal{A}$ iff $\frac{d\nu_1}{d\mu} \le \frac{d\nu_2}{d\mu}$ μ-a.e.

(3) If $\nu_n \lll \mu$ $\forall n$ & $\nu_n(A) \uparrow$ $\forall A \in \mathcal{A}$ then $\nu(\cdot) := \lim_n \nu_n(\cdot) \lll \mu$ and

$$\frac{d\nu_n}{d\mu} \xrightarrow{\mu\text{-a.e.}} \frac{d\nu}{d\mu}$$

11

(4) If $\nu \lll \mu$ then

ν is finite $\iff \frac{d\nu}{d\mu} \in L_1(\Omega, \mathcal{A}, \mu)$ and

ν is σ-finite $\iff \frac{d\nu}{d\mu} < \infty$ μ-a.e.

(5) If $\nu \lll \sigma \lll \mu$ then $\nu \lll \mu$ and

$$\frac{d\nu}{d\mu} = \frac{d\nu}{d\sigma}\frac{d\sigma}{d\mu} \quad \mu\text{-a.e.}$$

and $$\frac{d\nu}{d\sigma} = \frac{d\nu/d\mu}{d\sigma/d\mu} I_{\{\frac{d\sigma}{d\mu} > 0\}} \quad \sigma\text{-a.e.}$$

(6) if both $\mu \lll \nu$ & $\nu \lll \mu$ then

$\frac{d\nu}{d\mu} > 0$ μ-a.e and

$$\frac{d\mu}{d\nu} = \frac{1}{d\nu/d\mu} \quad \nu\text{-a.e. \& } \mu\text{-a.e.}$$

Proof:

For (1): Just check RHS integrates correctly.

For (2): Since μ is σ-finite our results on indefinite integrals applies & can be re-stated to say $\int_\cdot \frac{d\nu_1}{d\mu} d\mu \le \int_\cdot \frac{d\nu_2}{d\mu} d\mu$ $\forall A \in \mathcal{A}$

$\Updownarrow$ Lecture 11

$$\frac{d\nu_1}{d\mu} \le \frac{d\nu_2}{d\mu} \quad \mu\text{-a.e.}$$

This proves (2).

12

For (3): First note that $\nu(A)$ is defined by monotonicity & is a measure since clearly $\nu(\emptyset) = 0$ and

$$\begin{aligned}
\nu(\underbrace{\textstyle\bigcup_p A_p}_{\text{disjoint}}) &= \lim_n \nu_n(\textstyle\bigcup_p A_p) \\
&= \lim_n \sum_p \nu_n(A_p) \\
&= \sum_p \lim_n \nu_n(A_p) \quad \text{by Monotone Convergence Thm (MCT)} \\
&= \sum_p \nu(A_p).
\end{aligned}$$

Clearly $\nu \lll \mu$ and by (2)

$$0 \le \frac{d\nu_n}{d\mu} \le \frac{d\nu_{n+1}}{d\mu} \quad \mu\text{-a.e}$$

$$\begin{aligned}
\therefore \nu(A) &:= \lim_n \nu_n(A) \\
&= \lim_n \int_A \frac{d\nu_n}{d\mu} d\mu \\
&= \int_A \underbrace{\lim_n \frac{d\nu_n}{d\mu}}_{\text{exists by monotonicity}} d\mu \quad \text{by MCT}
\end{aligned}$$

For (4)-(6): These follow from our old results on densities, HWK 7 from Stat 235A, and similar arguments to the RND Thm.

QED

Lebesgue Decomposition

In some sense the Lebesgue decomposition result is a tweak to the RND Thm that allows us to say something when $P \not\ll Q$. We need it for studying likelihood ratios with martingales.

Theorem: (Lebesgue Decomposition)

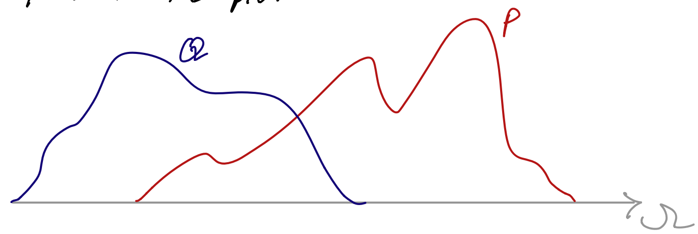
Let P and Q be two probability measures on $(\Omega, \mathcal{A})$. Then there exists two measures $Q_{\ll}$, $Q_{\perp}$ on $(\Omega, \mathcal{A})$ s.t.

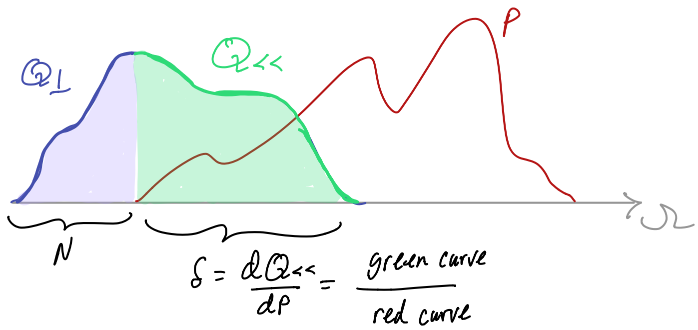
$$Q = \underbrace{Q_{\ll}}_{\substack{\text{where this is the} \\ P\text{-largest measure} \leq Q \\ \text{that is } \ll P}} + \underbrace{Q_{\perp}}_{\substack{\text{where } Q_{\perp}(\cdot) = Q(\cdot \cap N) \\ \text{for } P(N)=0. \\ \therefore Q_{\perp} \perp P}} \qquad (*)$$

Remark: $Q_{\ll}$ is the P-largest measure $\leq Q$ that is $\ll P$ means that for any other $\tilde{Q}$ measure which satisfies $\tilde{Q}(\cdot) \leq Q(\cdot)$ & $\tilde{Q} \ll P$ then it must be the case that

$$\frac{d\tilde{Q}}{dP} \leq \frac{dQ_{\ll}}{dP} \quad P\text{-a.e.}$$

Here is the picture:





Proof:

Recall in the proof of the first RND Thm to show $\frac{dQ}{dP}$ existed we set

$W = \frac{P+Q}{2}$ and showed

Found by Riesz

$$\frac{dQ}{dP} = \frac{dQ/dW}{dP/dW} I_{\{dP/dW \neq 0\}}$$

$$\overset{\substack{P\text{-a.e.} \\ Q\text{-a.e.}}}{=} \frac{dQ/dW}{dP/dW} \quad \text{by } P\left(\tfrac{dP}{dW}=0\right)=0 \overset{Q \ll P}{\Longrightarrow} Q\left(\tfrac{dP}{dW}=0\right)=0$$

But for this Thm we don't have $Q \ll P$ so lets simply define

$$Q_{\ll}(\cdot) = \int_{\cdot} \frac{dQ/dW}{dP/dW} I_{\{dP/dW \neq 0\}} \, dP$$

where dQ/dW & dP/dW exists from the RND thm by the fact that $P, Q \ll W$.

Now $\forall A \in \mathcal{Q}$

$$Q(A) = \underbrace{Q\left(A \cap \left\{\tfrac{dP}{dw} \neq 0\right\}\right)}_{} + \underbrace{Q\left(A \cap \left\{\tfrac{dP}{dw} = 0\right\}\right)}_{=: Q_\perp(A)}$$

$$= \int_A I_{\{\frac{dP}{dw} \neq 0\}} \frac{dQ}{dw}\, dw$$

$=: Q_\perp(A)$ which is $\perp$ to P since $\{\frac{dP}{dw} = 0\}$ is P-null.

$$= \int_A I_{\{\frac{dP}{dw} \neq 0\}} \frac{dQ/dw}{dP/dw} \frac{dP}{\cancel{dw}} \cancel{dw}$$

$$= Q_{<<}(A)$$

This proves $Q = Q_{<<} + Q_\perp$ where $Q_{<<} \ll P$ & $Q_\perp \perp P$

To show $Q_{<<}$ is P-largest let $\tilde{Q}$ be a measure s.t. $\tilde{Q}(\cdot) \leq Q(\cdot)$ & $\tilde{Q} << P$. Now let $N = \{\frac{dP}{dw} = 0\}$ and notice

$$\int_A \frac{d\tilde{Q}}{dP} dP = \int_{A \cap N} \frac{d\tilde{Q}}{dP} dP + \int_{A \cap N^c} \frac{d\tilde{Q}}{dP} dP$$

$$\leq \int_{A \cap N} \frac{d\tilde{Q}}{dP} dP + Q(A \cap N^c)$$

since $\tilde{Q}(\cdot) \leq Q(\cdot)$

$$= \underbrace{0}_{\text{since } P(N)=0} + \underbrace{Q_{<<}(A \cap N^c)}_{\text{since } Q_\perp(A\cap N^c) = Q(A \cap N^c \cap N) = 0}$$

$$= \int_{A \cap N^c} \frac{dQ_{<<}}{dP} dP$$

$$= \int_A I_{N^c} \frac{dQ_{<<}}{dP} dP$$

but this is 1 P-a.e. since $P(N)=0$

$$\therefore \int_A \frac{d\tilde{Q}}{dP} dP \leq \int_A \frac{dQ_{<<}}{dP} dP \quad \forall A \in \mathcal{Q}$$

$\therefore \frac{d\tilde{Q}}{dP} \leq \frac{dQ_{<<}}{dP}$ P-a.e. by our result on indefinite integrals in Lecture 11.

QED

Example

$$(\Omega, \mathcal{Q}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

$$dP = e^{-x} I_{(0,\infty)}(x)\, dx$$

$$dQ = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

Lets find $Q_{<<}$ & $Q_\perp(\cdot) = Q(\cdot \cap N)$. ($N$: P-Null.)

Set $N = [-\infty, 0]$ which is P-null and

$$\delta(x) = \begin{cases} e^{x} e^{-x^2/2} (2\pi)^{-1/2} & \text{for } x > 0 \\ 0 & \text{o.w.} \end{cases}$$

Now

$$\int_A \delta\, dP + Q(A \cap N)$$

$$= \int_A \delta(x) e^{-x} I_{(0,\infty)}(x)\, dx + Q(A \cap N)$$

by step in the density

$$= \int_{A \cap N^c} e^{-x^2/2} (2\pi)^{-1/2} dx + Q(A \cap N)$$

$$= Q(A \cap N^c) + Q(A \cap N)$$

$$= Q(A)$$

$\therefore Q_{<<}(\cdot) = \int_{\cdot} \delta\, dP$ & $Q_\perp(\cdot) = Q(\cdot \cap N)$ satisfies the LD Thm.

(Note the LD is unique but we only need that $Q_{<<}$ is P-largest later)