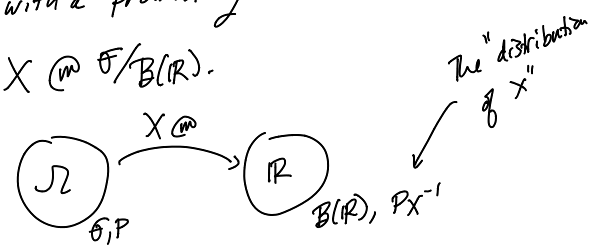
Lecture 11: Random variable densities and expected values

Recall that random variable is a map $X : \Omega \to \mathbb{R}$ along with a probability measure $(\Omega, \mathcal{F}, P)$ s.t. X ⓜ $\mathcal{F}/\mathcal{B}(\mathbb{R})$.



Def: The expected value of X, denoted $E(X)$, is defined as

$$E(X) = \int_\Omega X\,dP = \int_\Omega X^+ dP - \int_\Omega X^- dP$$

when it is defined, i.e. when $X \in \mathcal{Q}(\Omega, \mathcal{F}, P)$.

You should think of X as a placeholder for a random number $X(\omega)$ obtained by choosing $\omega \in \Omega$ at random according to P. Then $E(X)$ is essentially what you "expect" X to be.

e.g. For $\theta \in [0,1]$ the r.v. X has a Bernolli θ distribution, denoted $X \sim Ber(\theta)$, if

$$P(X = 1) = \theta$$

$$P(X = 0) = 1 - \theta.$$

Let $A = \{\omega : X(\omega) = 1\}$ so that

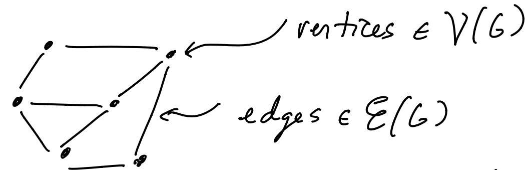
$$X(\omega) = I_A(\omega) \quad P\text{-a.e.}$$

$$\therefore\ E(X) = E(I_A(\omega)) \quad \text{by "a.e. useful"}$$
$$= P(A) \quad \text{by def.}$$
$$= \theta \quad \text{since } A = \{X = 1\}.$$

Note: It is always the case that

$$E(I_A) = P(A) \text{ whenever } A \in \mathcal{F}.$$

e.g. This example uses expected value & probability to prove a "non-probabilistic" statement about graphs G:



These types of proofs were made famous by Erdös.

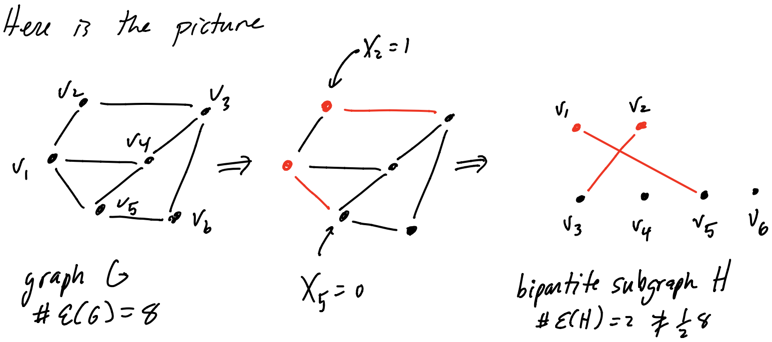
Claim: Every graph G has a bipartite subgraph H for which $\#\mathcal{E}(H) \geq \frac{1}{2}\#\mathcal{E}(G)$.

Proof:
Suppose G has n vertices labeled $v_1, v_2, \ldots, v_n$. Let $X_1, X_2, \ldots, X_n$ be n independent $Ber(\frac{1}{2})$ r.v.s defined on some probability space $(\Omega, \mathcal{F}, P)$. Define the subgraph H as follows

$$\mathcal{V}(H) := \mathcal{V}(G)$$

$$\mathcal{E}(H) := \{v_i v_j \in \mathcal{E}(G) : (X_i, X_j) = (1,0) \text{ or } (X_i, X_j) = (0,1)\}$$

Here is the picture



Notice that $\#\mathcal{E}(H)$ is a r.v. in $\mathcal{M}_s(\Omega, \mathcal{F})$. Let $\mathcal{I} = \{(i,j) : i > j \text{ \& } v_i v_j \in \mathcal{E}(G)\}$ index all edges $\mathcal{E}(G)$ so that

Now

$$E(\#\mathcal{E}(H)) = E\Big(\sum_{(i,j)\in\mathcal{X}} I_{\{(X_i,X_j)=(1,0)\}} + I_{\{(X_i,X_j)=(0,1)\}}\Big)$$

$$\overset{\text{Big 3}}{=} \sum_{(i,j)\in\mathcal{X}} P(X_i=1, X_j=0) + P(X_i=0, X_j=1)$$

$$= \sum_{(i,j)\in\mathcal{X}} \left(\tfrac{1}{2}\cdot\tfrac{1}{2} + \tfrac{1}{2}\cdot\tfrac{1}{2}\right)$$

$$= \frac{1}{2} \underbrace{\sum_{(i,j)\in\mathcal{X}} 1}_{\#\mathcal{E}(G)} \qquad (*)$$

Now if $\#\mathcal{E}(\tilde{H}) < \frac{1}{2}\mathcal{E}(G)$ for all bipartite subgraphs $\tilde{H}$ of G then

$$E(\#\mathcal{E}(H)) < \tfrac{1}{2}\#\mathcal{E}(G)$$

easy to see since $\#\mathcal{E}(H)$ is a simple r.v.

This contradicts $(*)$ so that we must have $\exists$ a bipartite subgraph $\tilde{H}$ s.t.

$$\#\mathcal{E}(\tilde{H}) \geq \tfrac{1}{2}\#\mathcal{E}(G).$$

QED

Lets look at a couple fundamental properties of expected value.

Theorem (Jensens inequality)

If $\varphi: \mathbb{R}\to\mathbb{R}$ is a convex function and $X \in L_1(\Omega,\mathcal{F},P)$ then $\varphi(X)$ is quasi-integrable and

$$\varphi(E(X)) \leq E(\varphi(X)).$$

(It is one of the great tragedies of probability & stats that this isn't an equality.)

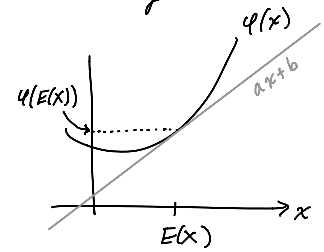
Proof:

Let $x \mapsto ax+b$ be a supporting line of φ passing through the point $(\underset{\sim}{E(X)}, \varphi(\underset{\sim}{E(X)}))$

finite.

In particular let $a,b\in\mathbb{R}$ satisfy

$$\begin{cases} ax+b \leq \varphi(x) \;\; \forall x\in\mathbb{R} \\ a\,E(X)+b = \varphi(E(X)) \end{cases}$$



$\therefore\; aX+b \leq \varphi(X) \quad (*)$

Notice that $aX+b$ is integrable since X is integrable.

Also φ is convex & mapping into $\mathbb{R}$ $\Rightarrow$ φ is continuous

$\Rightarrow$ φ ⓜ

$\Rightarrow$ $\varphi(X)$ is a r.v. and in $\mathcal{Q}^-$ by $(*)$.

$$\therefore\; \underbrace{a\,E(X)+b}_{\substack{\| \\ \varphi(E(X))}} \leq E(\varphi(X)) \quad \text{by Big 3}$$

QED

e.g. Lets use Jensen's inequality to get an understanding of why type of maximal fluctuation we might expect to see if we stare at a large number of random variables

Theorem (what to expect of the max)

Let $X_1, X_2, \ldots, X_n$ be r.v.s in $L_1(\Omega,\mathcal{F},P)$.

Suppose $\exists\sigma>0$ s.t.

$$E(e^{tX_i}) \leq \exp\left(\tfrac{t^2\sigma^2}{2}\right), \quad \forall t>0, \forall i\leq n$$

Then $E\left(\max_{1\leq i\leq n} X_i\right) \leq \sigma\sqrt{2\log n}.$

Proof:

Jensen

$$\exp\left(t E\left(\max_{i\le n} X_i\right)\right) \le E\left(\exp\left(t \max_{i\le n} X_i\right)\right)$$

$$= E\left(\max_{i\le n} \exp(tX_i)\right)$$

since $e^{tx} \uparrow$ in x

$$\le E\left(\sum_{i\le n} \exp(tX_i)\right)$$

since these are ≥ 0

Big3

$$= \sum_{i\le n} E(\exp(tX_i))$$

$$\le n\exp\left(\frac{t^2\sigma^2}{2}\right) \text{ by assumption.}$$

Taking log gives

$$E\left(\max_{i\le n} X_i\right) \le \underbrace{\frac{\log n}{t} + \frac{t\sigma^2}{2}}_{\text{Now choose a good } t} \quad \forall t>0.$$

Notice

$$\frac{d}{dt}\left(\frac{\log n}{t} + \frac{t\sigma^2}{2}\right) = -\frac{\log n}{t^2} + \frac{\sigma^2}{2} = 0$$

$$\Updownarrow$$

$$t = \frac{\sqrt{2\log n}}{\sigma}$$

Plugging this into our inequality gives

$$E\left(\max_{i\le n} X_i\right) \le \frac{\sigma}{\sqrt{2\log n}}\log n + \frac{\sqrt{2\log n}}{\sigma}\frac{\sigma^2}{2}$$

$$= \frac{\sigma}{\sqrt{2}}\sqrt{\log n} + \frac{\sigma}{\sqrt{2}}\sqrt{\log n}$$

$$= \sigma\sqrt{2\log n} \qquad \text{QED}$$

Theorem (expected value factors on indep r.v.s)

Suppose X and Y are (possibly extended) independent r.v.s on $(\Omega, \mathcal{F}, P)$. If $X \ge 0$ & $Y \ge 0$ or $X, Y \in L_1(\Omega, \mathcal{F}, P)$ then $XY \in Q(\Omega, \mathcal{F}, P)$ and

$$E(XY) = E(X)E(Y).$$

Proof:

Case 1: Suppose $X, Y \in \mathcal{M}_s(\Omega, \mathcal{F}, P)$ so that

$$X = \sum_{i=1}^{n} a_i I_{A_i} \;\&\; Y = \sum_{j=1}^{m} b_j I_{B_j}$$

where $a_1, \dots, a_n$ are distinct and $A_1, \dots, A_n$ are a disjoint measurable partition of Ω (and same for b_j's & B_j's).

Note that A_i is indep of B_j since

$$\left.\begin{aligned} A_i &= \{X = a_i\} \in \sigma\langle X\rangle \\ B_j &= \{Y = b_j\} \in \sigma\langle Y\rangle \end{aligned}\right\} \text{indep } \sigma\text{-fields}$$

$$\therefore\; E(\underbrace{XY}_{\in \mathcal{M}_s}) = E\left(\sum_{i,j} a_i b_j I_{A_i \cap B_j}\right)$$

$$= \sum_{i,j} a_i b_j \underbrace{P(A_i \cap B_j)}_{P(A_i)P(B_j)}$$

Note that a_i or b_j could be ∞ but $I_{A_i \cap B_j} \in \mathcal{M}$ allows Big 3(2)[a] to apply

$$= \left(\sum_i a_i P(A_i)\right)\left(\sum_j b_j P(B_j)\right)$$

$$= E(X)E(Y).$$

Case 2: Suppose $X, Y \in \mathcal{M}(\Omega, \mathcal{F})$.

Notice that we also have that

$$X \in \mathcal{M}(\Omega, \sigma\langle X\rangle) \;\&\; Y \in \mathcal{M}(\Omega, \sigma\langle Y\rangle).$$

Therefore the structure theorem implies there exists $X_n \in \mathcal{M}_s(\Omega, \sigma\langle X\rangle)$ and $Y_n \in \mathcal{M}_s(\Omega, \sigma\langle Y\rangle)$ such that

$$X_n \uparrow X \text{ and } Y_n \uparrow Y.$$

$\therefore \quad E(X_nY_n) = E(X_n)E(Y_n)$ by case 1.

$$\begin{aligned}\therefore\ E(XY) &= E(\lim_n\uparrow X_nY_n)\\ &= \lim_n\uparrow E(X_nY_n) \text{ by little 3}\\ &= \lim_n\uparrow E(X_n)E(Y_n)\\ &= E(X)E(Y).\end{aligned}$$

Case 3: Suppose $X, Y \in L_1(\Omega, \mathcal{F}, P)$.

Notice that

$$(XY)^+ = X^+Y^+ + X^-Y^-$$
$$(XY)^- = X^+Y^- + X^-Y^+$$

Therefore ... all finite

$$E(XY)^+ = E(X^+)E(Y^+) + E(X^-)E(Y^-) < \infty$$
$$E(XY)^- = E(X^+)E(Y^-) + E(X^-)E(Y^+) < \infty$$

by little 3, case 2 and the fact that $\sigma\langle X^+\rangle \subset \sigma\langle X\rangle, \ldots$ & $\sigma\langle Y^-\rangle \subset \sigma\langle Y\rangle$.

(Notice I'm implicitly using "X @ w.r.t $\sigma\langle Y\rangle$ thm" & "subclasses")

$\therefore\ XY \in L_1(\Omega, \mathcal{F}, P)$ and

$$\begin{aligned}E(XY) &= E(XY)^+ - E(XY)^-\\ &= E(X^+)E(Y^+) + E(X^-)E(Y^-)\\ &\quad - E(X^+)E(Y^-) - E(X^-)E(Y^+)\\ &= E(X^+)E(Y) - E(X^-)E(Y)\\ &= E(XY).\end{aligned}$$

QED

Notice this fully generalizes to situations like this: Suppose $X_1, X_2, \ldots$ are independent $L_1(\Omega, \mathcal{F}, P)$ r.v.s then

$$E(X_1X_2\cdots X_n) = E(X_1)E(X_2)\cdots E(X_n)$$

and

$$E\big(g(X_1, X_3, \ldots)h(X_2, X_4, \ldots)\big) = Eg(X_1, X_3, \ldots)\, Eh(X_2, X_4, \ldots)$$

if $g, h \in \mathcal{M}(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$. To show these just use the previous Thm, ANOVA & "X @ w.r.t $\sigma\langle Y\rangle$ Thm".

Def: If X and Y are two r.v.s on $(\Omega, \mathcal{F}, P)$ then set

$$var(X) \equiv \text{"the variance of } X\text{"} := E(X - E(X))^2$$

$$sd(X) \equiv \text{"the standard deviation of } X\text{"} := \sqrt{var(X)}$$

$$cov(X,Y) \equiv \text{"the covariance btwn } X \,\&\, Y\text{"} =: E\big[(X - E(X))(Y - E(Y))\big].$$

when they are defined.

Note: $var(X)$, $sd(X)$ & $cov(X,Y)$ may not be defined if the expectations which define them do not exist.

Remark: $sd(X)$ measures how spread out X is on $\mathbb{R}$ & $cov(X,Y)$ measure how X & Y co-vary together.

Later we will see that $sd(X)$ and $cov(X,Y)$ are essentially the functional analysis equivalent to L_2 norm & L_2 inner product. Here is a hint as to why

Theorem (Hölder)

Let X and Y be two R.V.s on $(\Omega, \mathcal{F}, P)$. If $p,q>0$ s.t. $\frac{1}{p}+\frac{1}{q}=1$ then

$$E|XY| \le \left(E|X|^p\right)^{1/p}\left(E|Y|^q\right)^{1/q} \qquad (*)$$

even if X and Y are not quasi-integrable. If $XY \in Q(\Omega,\mathcal{F},P)$ then

$$|E(XY)| \le \left(E|X|^p\right)^{1/p}\left(E|Y|^q\right)^{1/q}. \qquad (**)$$

Proof: $(*)$ is trivially true if any one of the following is true:

$$E|X|^p=\infty,\ E|X|^p=0,\ E|X|^q=\infty,\ \underbrace{E|X|^q=0}.$$

since "a.e. useful facts" implies $|X|^q \overset{a.e.}{=} 0$ so that $|XY| \overset{a.e.}{=} 0$.

So suppose $E|X|^p, E|Y|^q \in (0,\infty)$.

Define $Z := \dfrac{X}{(E|X|^p)^{1/p}}$ & $W := \dfrac{Y}{(E|Y|^q)^{1/q}}$

Now we simply show

$$E|ZW| \le 1.$$

Use Young's inequality

$$a^{w_1}b^{w_2} \le w_1 a + w_2 b \qquad (***)$$

when $a,b\ge 0$ & $w_1,w_2>0$ s.t. $w_1+w_2=1$

Young's inequality follows since log is concave so that

$$w_1\log a + w_2 \log b \le \log(w_1 a + w_2 b).$$

Now $E|ZW| = E\Big(\underbrace{(|Z|^p)^{1/p}(|W|^q)^{1/q}}\Big)$

has the form $(***)$ when $a=|Z(\omega)|^p$, $b=|W(\omega)|^q$, $w_1=\frac{1}{p}$ & $w_2=\frac{1}{q}$.

$$\overset{(***)}{\le} \frac{1}{p}\underbrace{E|Z|^p}_{=1}+\frac{1}{q}\underbrace{E|W|^q}_{=1}, \text{ using Big 3(2)[1]}$$

$$=1.$$

If $XY \in Q(\Omega,\mathcal{F},P)$ then

$$|E(XY)| \le E|XY|$$

by corollary to Big 3.

QED.

Corollary:

If X and Y are two r.v.s on $(\Omega,\mathcal{F},P)$ s.t. $E(X^2)<\infty$ & $E(Y^2)<\infty$ then $cov(X,Y)$, $sd(X)$ & $sd(Y)$ are well defined and

$$|cov(X,Y)| \le sd(X)\,sd(Y)$$

Proof: By Hölder $E|X| \le \sqrt{E|X|^2} < \infty$ so that $X,Y \in L_1(\Omega,\mathcal{F},P)$ and $E(X)<\infty$, $E(Y)<\infty$. Let $\tilde{X}=X-E(X)$, $\tilde{Y}=Y-E(Y)$. Also by Hölder we have $E|\tilde{X}\tilde{Y}| \le \sqrt{E(\tilde{X}^2)}\sqrt{E(\tilde{Y}^2)} < \infty$

$\therefore\ \tilde{X}\tilde{Y} \in Q$ & $|cov(X,Y)| \le \sqrt{E(\tilde{X}^2)}\sqrt{E(\tilde{Y}^2)}$

$$= sd(X)\,sd(Y)$$

QED

Corollary:

If X and Y are two independent r.v.s s.t. $E(X^2)<\infty$ & $E(Y^2)<\infty$ then $cov(X,Y)$, $E(X)$ and $E(Y)$ are well defined, finite and

$$cov(X,Y) = E\big[(X-EX)(Y-EY)\big]$$
$$= \underbrace{E(X-EX)}_{=0}\,\underbrace{E(Y-EY)}_{=0}$$

Once we prove the following theorem we will discuss how it gives an interesting relation btwn Lebesgue integration and Riemann integration.

Theorem:

Suppose X is a r.v. on $(\Omega,\mathcal{F},P)$ s.t. $X\geq 0$ P-a.e.. Then

$$E(X) \overset{(*)}{=} \int_0^\infty P(X>t)\,dt \overset{(**)}{=} \int_0^\infty P(X\geq t)\,dt$$

Lebesgue integral

Proof:

First notice that $t\mapsto P(X\geq t)$ & $t\mapsto P(X>t)$ are both monotonic and therefore ⓜ $\mathcal{B}(\mathbb{R})/\mathcal{B}(\mathbb{R})$.

Clearly these two functions are in $\mathcal{B}^-([0,\infty], \mathcal{B}([0,\infty]), \mathcal{L}^1)$ so the integrals in $(*)$ and $(**)$ are well defined.

To show $(*)$ use the "1-2 argument".

Step 1: Show $(*)$ holds for $X\in\mathcal{M}_s(\Omega,\mathcal{F})$

$\therefore$ X can be written in the form

$$X = \sum_{i=1}^n \underbrace{c_i}_{\in[0,\infty)} I_{A_i}$$

measurable partition

So that $E(X) = \sum_{i=1}^n c_i P(A_i)$.

Now

$$\int_0^\infty P(X>t)\,dt = \int_0^\infty \sum_{i=1}^n P(\{X>t\}\cap A_i)\,dt$$

since A_i's partition Ω

$$\overset{\text{Big 3}}{=} \sum_{i=1}^n \int_0^\infty P(\{X>t\}\cap A_i)\,dt$$

on A_i, $X=c_i$

$$= \sum_{i=1}^n \int_0^\infty \underbrace{P(\{c_i>t\}\cap A_i)}_{=\begin{cases}0 & \text{if } c_i\leq t\\ P(A_i) & \text{o.w.}\end{cases}}\,dt$$

$$= \sum_{i=1}^n c_i P(A_i)$$

$$= E(X)$$

Step 2:

If $X\in\mathcal{M}(\Omega,\mathcal{F})$ then $\exists X_n\in\mathcal{M}_s(\Omega,\mathcal{F})$ s.t.

$$X_n \uparrow X.$$

$$\therefore\ E(X) = E(\lim_n\uparrow X_n)$$
$$= \lim_n\uparrow E(X_n), \quad \text{little 3}$$
$$= \lim_n\uparrow \int_0^\infty P(X_n>t)\,dt$$
$$= \int_0^\infty \lim_n\uparrow P(X_n>t)\,dt, \quad \text{little 3}$$
$$= \int_0^\infty P(X>t)\,dt$$

by CFB since $\{X_n>t\}\uparrow\{X>t\}$ (but not $\{X_n\geq t\}\not\uparrow\{X\geq t\}$)

This gives $(*)$.

To show (**) simply notice that

$$P(X \geq t) = P(X > t) + \underbrace{P(X = t)}_{\substack{\text{this can only be non-zero} \\ \text{for at most countably} \\ \text{many } t\text{'s.}}}$$

$$= P(X > t) \quad \mathcal{L}^1\text{-a.e.}$$

QED

The previous theorem can be useful for computing the expected value of the minimum of a sequence of r.v.

e.g. Let $U_1, U_2, \ldots, U_n$ be iid $Unif((0,1))$ random variables. Then

$$E\Big(\min_{1 \le i \le n} U_i\Big) = \int_0^\infty P\Big(\underbrace{\min_{1 \le i \le n} U_i \geq t}_{\substack{\text{This event holds} \\ \text{iff each } U_i \geq t.}}\Big)dt$$

$$= \int_0^\infty P(U_1 \geq t, \ldots, U_n \geq t)\,dt$$

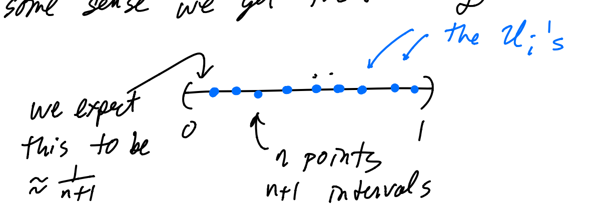
$$= \int_0^\infty P(U_1 \geq t)^n\,dt$$

since U_i's are indep and identically distributed

$$= \int_0^1 (1-t)^n\,dt$$

$$= -\frac{(1-t)^{n+1}}{n+1}\Bigg|_0^1 = \frac{1}{n+1}$$

In some sense we get the following picture



Using improper Riemann integration to define $E(X)$.

To relate the formula $E(X) = \int_0^\infty P(X > t)\,dt$ to Riemann integration first notice

Thm (cdf's have countably many jumps)

If X is a r.v. with cdf $F(t) := P(X \le t)$ then $\{t \in \mathbb{R} : F \text{ is discontinuous } @t\}$ is countable

Proof:

Since $F(t)$ is right continuous

$$F \text{ is discontinuous } @t \iff F(t) \neq F(t-)$$
$$\iff F(t) - F(t-) > 0$$
$$\iff P(X = t) > 0$$

But $\{\{X = t\} : t \in \mathbb{R}\}$ is a countable collection of disjoint events $\Longrightarrow P(X=t) > 0$ for at most countably many $t \in \mathbb{R}$ (by a Thm in Lecture 5).

QED.

This means that $t \mapsto P(X > t) = P(X \le t)$ is Riemann integrable on any bdd interval $t \in [a,b]$ (since it is bdd and has countably many discontinuities)

Now when $X \geq 0$

$$E(X) = \int_0^\infty P(X > t)\,dt$$

abstract integral $\int_\Omega X(\omega)\,dP(\omega)$

This can be interpreted as a improper Riemann integral on $[0,\infty]$.
... this can be used to define $E(X)$ with just improper Riemann integration!

Probability inequalities involving expected value

It is often the case that computing bounds for expected value is easy. These inequalities then yeild probability bounds

Theorem (Markov's inequality)

If X is a r.v. on $(\Omega, \mathcal{F}, P)$ s.t. $X \geq 0$ P-a.e then $P(X \geq \alpha) \leq \frac{E(X)}{\alpha}$ for all $\alpha > 0$.

Proof: $X \geq 0$ P-a.e. implies that $X \in Q^-(\Omega, \mathcal{F}, P)$ and $X I_{\{X \geq \alpha\}} \geq \alpha I_{\{X \geq \alpha\}}$ "Indicate what you want trick"

$$\begin{aligned} \therefore \quad E(X) &\geq E(X I_{\{X \geq \alpha\}}) && \text{by Big 3(i) and } X \geq 0 \text{ P.a.e.} \\ &\geq E(\alpha I_{\{X \geq \alpha\}}) && \text{by Big 3(i)} \\ &= \alpha P(X \geq \alpha) && \text{definition of } \textstyle\int_\Omega dP \end{aligned}$$

QED.

Corollary:

For any X on $(\Omega, \mathcal{F}, P)$ and any $\alpha > 0$:

i) $P(|X| \geq \alpha) \leq \frac{E(|X|^k)}{\alpha^k}$ could be ∞

ii) $P(X \geq \alpha) \leq \inf_{t>0} \frac{E(e^{tX})}{e^{t\alpha}}$ chernoffs method

iii) if $E(X^2) < \infty$ then $E(X) < \infty$ and $P(|X - EX| \geq \alpha) \leq \frac{var(X)}{\alpha^2}$ Chebyshev's inequality

Note: i) says that if $E|X|^k < \infty$ for a large $k > 0$ then $P(|X| \geq \alpha)$ decays quickly as $\alpha \to \infty$.
we used ii) in Lecture 1 when illustrating the "modern" way to prove Borel's Normal number theorem.
iii) shows why $sd(X)$ controls the "spread" of X.

Lets use Chebyshev to show a famous theorem in analysis.
e.g

Theorem (Weierstrauss Approximation)

If $f : [0,1] \to \mathbb{R}$ is continuous then for any $\varepsilon > 0$ there exists a polynomial $p(x)$ s.t.

$$\sup_{x \in [0,1]} |f(x) - p(x)| < \varepsilon.$$

Proof: Let $U_1, U_2, \ldots$ be iid r.v.s uniformly distributed on $(0,1)$. For each $x \in [0,1]$ let $F_x(\cdot)$ be the c.d.f of a $Ber(x)$ r.v. and set

$$S_n^x := \underbrace{F_x^{-1}(U_1) + \cdots + F_x^{-1}(U_n)}_{\text{independent } Ber(x) \text{ r.v.s}}$$

Note that S_n^x is a collection of coupled r.v.s indexed by x

A simple counting argument shows

$$P(S_n^x = m) = \binom{n}{m} x^m (1-x)^{n-m}$$

$S_n^x \sim Bin(x, n)$

for $m = 0, 1, \ldots, n$. Moreover,

$$E\left(\frac{S_n^x}{n}\right) = x \quad \text{and} \quad var\left(\frac{S_n^x}{n}\right) = \frac{x(1-x)}{n}.$$

Also notice that

$$P_n(x) := E \underbrace{f\left(\frac{S_n^x}{n}\right)}_{\text{simple r.v.}} = \underbrace{\sum_{m=0}^{n} f\left(\frac{m}{n}\right) P(S_n^x = m)}_{\text{Polynomial in } x \text{ of degree } n.}$$

Since f is uniformly continuous on $[0,1]$ let

$$M := \sup_{x \in [0,1]} |f(x)| < \infty$$

$$\delta(\varepsilon) := \sup\{ |f(x) - f(y)| : |x - y| < \varepsilon \}$$

and for a given $\varepsilon > 0$ and $x \in [0,1]$ let

$$A_{\varepsilon, x} := \left\{ \omega \in \Omega : \left| \frac{S_n^x(\omega)}{n} - x \right| < \varepsilon \right\}$$

The definition of $A_{\varepsilon,x}$ now implies

$$\omega \in A_{\varepsilon,x} \Rightarrow \left| f\left(\frac{S_n^x(\omega)}{n} \right) - f(x) \right| \le \delta(\varepsilon)$$

$$\omega \in A_{\varepsilon,x}^c \Rightarrow \left| f\left(\frac{S_n^x(\omega)}{n} \right) - f(x) \right| \le 2M$$

so that

$$\left| f\left(\frac{S_n^x(\omega)}{n} \right) - f(x) \right| = \left| f\left(\frac{S_n^x(\omega)}{n} \right) - f(x) \right| \left(I_{A_{\varepsilon,x}}(\omega) + I_{A_{\varepsilon,x}^c}(\omega) \right)$$
$$\le \delta(\varepsilon) I_{A_{\varepsilon,x}}(\omega) + 2M I_{A_{\varepsilon,x}^c}(\omega)$$

To finish

$$|p_n(x) - f(x)| = \left| E f\left(\frac{S_n^x}{n} \right) - f(x) \right|$$
$$\le E \left| f\left(\frac{S_n^x}{n} \right) - f(x) \right| \quad \text{by Jensen}$$
$$\le \delta(\varepsilon) P\left(\left| \frac{S_n^x}{n} - x \right| < \varepsilon \right) + 2M P\left(\left| \frac{S_n^x}{n} - x \right| \ge \varepsilon \right)$$
$$\le \delta(\varepsilon) + 2M \frac{\operatorname{var}\left(\frac{S_n^x}{n} \right)}{\varepsilon^2} \quad \text{by Chebyshev}$$
$$= \delta(\varepsilon) + 2M \frac{x(1-x)}{n\varepsilon^2}$$

Since $x(1-x) \le \frac{1}{4}$ $\forall x \in [0,1]$ we have

$$\sup_{x \in [0,1]} |p_n(x) - f(x)| \le \delta(\varepsilon) + \frac{M}{2n\varepsilon^2} \qquad (*)$$

By replacing ε with $\frac{1}{n^{1/3}}$ (for example) and choosing n large enough I can make $(*)$ as small as I want (since $\delta(\varepsilon) \to 0$ as $\varepsilon \to 0$).

QED

Chernoff's method is especially useful since $t \mapsto E(e^{tX})$ is an important function called the moment Generating function. More on that later.

Lets look at a special case which will give us the SLLN for bdd r.v.s.

Theorem (Hoeffding's inequality)

Let $X_1, X_2, \ldots, X_n$ be iid r.v.s on $(\Omega, \mathcal{F}, P)$. If there exists finite real numbers $a < b$ s.t.

$$a \le X_i \le b$$

P-a.e. $\forall i = 1, \ldots, n$, then

$$P\left(\left| \frac{S_n}{n} - \mu \right| \ge \varepsilon \right) \le 2 \exp\left(-\frac{2n\varepsilon^2}{(b-a)^2} \right)$$

$\forall \varepsilon > 0$ where $S_n := X_1 + \cdots + X_n$ and $\mu := E(X_1)$.

Note: The iid assumption in Hoeffding can be relaxed and extended to martingales. We will cover this next quarter when we study dependence in random variables.

Note: The Hoeffding bound gives the exact same bound we derived by hand for our radamacher coin flip r.v.s $R_1, R_2, \ldots$ from lecture 1:

$$P\left(\left| \frac{S_n}{n} \right| \ge \varepsilon \right) \le 2e^{-n\varepsilon^2/2}$$

where $\frac{S_n}{n} = \frac{R_1 + \cdots + R_n}{n}$ and $-1 \le R_i \le 1$.

Lemma: If X is a r.v. on $(\Omega, \mathcal{F}, P)$ s.t. $\exists$ finite $a<b$ s.t. $a \le X \le b$ & $E(X)=0$

Then $E(e^{tX}) \le e^{t^2(b-a)^2/8}$ $\forall t \geq 0$.

Proof:

Let $W = \frac{X-a}{b-a}$ so that $0 \le W \le 1$ and $X = Wb + (1-W)a$. By convexity we have

$$e^{tX} \le We^{tb} + (1-W)e^{ta}$$

$$\therefore E(e^{tX}) \le E(W)e^{tb} + E(1-W)e^{ta}$$

$$= \frac{E(X)-a}{b-a} e^{tb} + \frac{b-E(X)}{b-a} e^{ta}$$

$$= -\frac{a}{b-a} e^{tb} + \frac{b}{b-a} e^{ta}$$

$$= (1-\theta)e^{-u\theta} + \theta e^{u(1-\theta)}$$

where $u = -(b-a)t$ & $\theta = \frac{b}{b-a}$.

Notice the assumption $E(X)=0$ implies $0 \in [a,b]$ so that $0 \le \theta \le 1$.

If we can show

$$(1-\theta)e^{-u\theta} + \theta e^{u(1-\theta)} \le e^{u^2/8} \quad (*)$$

$\forall u \in \mathbb{R}$ and $\forall \theta \in [0,1]$ we are done.

Taking log it is sufficient to show

$$\log\left((1-\theta)e^{-u\theta} + \theta e^{u(1-\theta)}\right) \le u^2/8$$

$$\|$$

$$\log\left(e^{-u\theta}\left[1-\theta + \theta e^{u}\right]\right)$$

$$\|$$

$$-u\theta + \log(1-\theta+\theta e^{u})$$

$$\|$$

$$K(u)$$

Now $K'(u) = -\theta + \frac{\theta e^{u}}{1-\theta+\theta e^{u}} = -\theta + \frac{\theta}{\theta + (1-\theta)e^{-u}}$

$$K''(u) = \frac{\theta(1-\theta)e^{-u}}{(\theta + (1-\theta)e^{-u})^2}$$

Now Taylor's thm gives

$$K(u) = K(0) + uK'(0) + \frac{u^2}{2}K''(u^*), \quad u^* \in [0,u]$$

$$= 0 + 0 + \frac{u^2}{2} \underbrace{\underbrace{\left(\frac{\theta}{\theta + (1-\theta)e^{-u^*}}\right)}_{\in[0,1]}\underbrace{\left(1 - \frac{\theta}{\theta+(1-\theta)e^{-u^*}}\right)}_{\in[0,1]}}_{\le \frac{1}{4}}$$

$\therefore K(u) \le u^2/8$

QED.

Proof of Hoeffding's inequality:

We can suppose w.l.g that $E(X_i) = \mu = 0$.

Now

$$P\left(\left|\frac{S_n}{n}\right| \geq \varepsilon\right) = P\left(\left\{\frac{S_n}{n} \geq \varepsilon\right\} \cup \left\{-\frac{S_n}{n} \geq \varepsilon\right\}\right)$$

$$= P\left(\frac{S_n}{n} \geq \varepsilon\right) + P\left(-\frac{S_n}{n} \geq \varepsilon\right)$$

using Chernoff's method for any $t>0$

$$P\left(\frac{S_n}{n} \geq \varepsilon\right) \le P\left(e^{tS_n} \geq e^{tn\varepsilon}\right)$$

$$\le \frac{E(e^{tS_n})}{e^{tn\varepsilon}} \quad \text{by Markov ineq.}$$

$$= e^{-tn\varepsilon} \prod_{i=1}^{n} \underbrace{E(e^{tX_i})}_{\le e^{t^2(b-a)^2/8} \text{ by lemma}} \quad \text{by indep}$$

$$\le \underbrace{e^{-tn\varepsilon} e^{nt^2(b-a)^2/8}}_{\text{minimized at } t = \frac{4\varepsilon}{(b-a)^2}}$$

$$\therefore P\left(\frac{S_n}{n} \geq \varepsilon\right) \le e^{-4n\varepsilon^2/(b-a)^2} e^{n4^2\varepsilon^2/[8(b-a)^2]}$$

$$= e^{-2n\varepsilon^2/(b-a)^2}$$

Similar arguments give the exact same upper bound for $P\left(-\frac{S_n}{n} \geq \varepsilon\right)$.

QED

Here is an example of the utility of Hoeffding.

e.g.

Theorem: (SLLN for bounded r.v.s)

Let $X_1, X_2, \ldots$ be iid r.v.s on $(\Omega, \mathcal{F}, P)$ s.t. $|X_i| \le c$ for some finite c. Then

$$\frac{S_n}{n} \xrightarrow{n\to\infty} E(X_1) \quad P\text{-a.e.}$$

where $S_n = X_1 + \cdots + X_n$.

Proof:

By Hoeffding's inequality

$$P\left(\left|\frac{S_n}{n} - E(X_1)\right| \ge \varepsilon\right) \le \underbrace{2\exp\left(-\frac{2n\varepsilon^2}{(2c)^2}\right)}_{\text{finitely summable over } n \text{ for each } \varepsilon>0} \quad \forall \varepsilon > 0.$$

$\therefore$ The First Borel-Cantelli lemma applies so that $P\left(\bigcap_{\varepsilon\in\mathbb{Q}^+} \left\{\left|\frac{S_n}{n} - E(X_1)\right| < \varepsilon \text{ a.a. } n\right\}\right) = 1$

$\therefore$ $P\left(\frac{S_n}{n} \to E(X_1)\right) = 1.$

QED.

Notice that this tells us we have the right definition for $E(X) := \int_\Omega X\,dP$, i.e. $E(X)$ tells use the long run average of independent samples of X (with probability 1).

After we cover densities, later in this lecture, we will discuss the Glavenko-Cantelli Theorem which is an important application of the SLLN for bdd r.v.s.

In the Law of the iterated log we needed lower bounds to get tight control on the behavior of $P(S_n \ge \alpha)$. Let's look at one example of a lower bound.

Theorem: (Paley-Zygmund ineq.)

If X is a non-negative r.v. s.t. $E(X^2) < \infty$ then

$$(1-\alpha)^2 \frac{(EX)^2}{E(X^2)} \le P(X \ge \alpha EX)$$

for all $\alpha \in (0,1)$.

Proof: First notice that

$$X = X I_{\{X < \alpha EX\}} + X I_{\{X \ge \alpha EX\}}.$$

Taking expected values gives

$$\begin{aligned} E(X) &\le \alpha E(X) + E(X I_{\{X\ge\alpha EX\}}) \\ &\le \alpha E(X) + \sqrt{E(X^2)}\sqrt{E(I^2_{\{X\ge\alpha EX\}})} \quad \text{by Hölder} \\ &= \alpha E(X) + \sqrt{E(X^2)}\sqrt{P(X\ge\alpha EX)} \end{aligned}$$

Now since $E(X), E(X^2)$ are both finite ($E(X)<\infty$ by Hölder) we can shuffle terms around to get the result.

QED.

Note: Jensen's inequality shows $(EX)^2 \le E(X^2)$

$\therefore$ the lower bound $(1-\alpha)^2 \frac{(EX)^2}{E(X^2)} < 1$

Maximal Inequalities

We used two maximal inequalities in lecture 7 which were custom to our coinflip model. In this section we state general versions and use them later (after we get the CLT) to establish Kolmogorov's 3 series theorem.

Theorem (Kolmogorov's Maximal inequality)

Let $X_1, \ldots, X_n$ be independent r.v.s s.t. $E(X_k^2) < \infty$ and $E(X_k) = 0$ $\forall k$. If $\alpha > 0$ then

$$P\left(\max_{1 \le k \le n} |S_k| \ge \alpha\right) \le \frac{1}{\alpha^2} \underbrace{var(S_n)}_{= E(S_n^2)}$$

where $S_n := X_1 + \cdots + X_n$.

Proof: The proof is exactly similar to the one we did for coin flips in Lecture 7. Lets do it again using our theory of integration.

Define $F_k := \{ |S_1| < \alpha, \ldots, |S_{k-1}| < \alpha, |S_k| \ge \alpha \}$.

$$E(S_n^2) = \int_\Omega S_n^2 \, dP$$

$$\ge \int_\Omega S_n^2 \sum_{k=1}^n I_{F_k} \, dP \quad \text{since } F_k\text{'s are disjoint so } \sum_k I_{F_k} \le 1.$$

$$= \sum_{k=1}^n \int_\Omega S_n^2 I_{F_k} \, dP \qquad S_n^2 = (S_n \pm S_k)^2$$

$$= \sum_{k=1}^n \int_\Omega [S_k^2 + 2 S_k (S_n - S_k) + \underbrace{(S_n - S_k)^2}_{\text{drop this}}] I_{F_k} \, dP$$

$$\ge \sum_{k=1}^n \int_\Omega S_k^2 I_{F_k} \, dP + 2 \int_\Omega S_k I_{F_k} (S_n - S_k) \, dP$$

Notice S_k ⓜ $\sigma\langle X_1, \ldots, X_k \rangle$ since S_k is a measurable function of $X_1, \ldots, X_k$.
Also I_{F_k} ⓜ $\sigma\langle X_1, \ldots, X_k \rangle$ since $F_k \in \sigma\langle X_1, \ldots, X_k \rangle$.
Therefore $S_k I_{F_k}$ ⓜ $\sigma\langle X_1, \ldots, X_k \rangle$ so that

$$\sigma\langle S_k I_{F_k} \rangle \subset \sigma\langle X_1, \ldots, X_k \rangle$$

and similarly

$$\sigma\langle S_n - S_k \rangle \subset \sigma\langle X_{k+1}, \ldots, X_n \rangle$$

Since the X_k's are indep we have

$$E(S_k I_{F_k} (S_n - S_k)) = E(S_k I_{F_k}) \underbrace{E(S_n - S_k)}_{= 0}$$

$$\therefore E(S_n^2) \ge \int_\Omega \sum_{k=1}^n S_k^2 I_{F_k} \, dP \quad \text{on } |S_k| \ge \alpha \text{ on } F_k$$

$$\ge \alpha^2 \int_\Omega \underbrace{\sum_{k=1}^n I_{F_k}}_{I_{\{\max_{1 \le k \le n} |S_k| \ge \alpha\}}} \, dP$$

$$\ge \alpha^2 P\left(\max_{1 \le k \le n} |S_k| \ge \alpha\right)$$

QED.

Here is Etemadi's ineq which can be used without the moment assumptions on X_k in Kolmogorov's max ineq.

Theorem (Etemadi's maximal inequality) ← doesn't require iid

Suppose $X_1, X_2, \ldots, X_n$ are independent r.v.s and $\alpha \ge 0$. Then

$$P\left(\max_{1 \le k \le n} |S_k| \ge 3\alpha\right) \le 3 \max_{1 \le k \le n} P(|S_k| \ge \alpha)$$

Proof: The proof is similar to what we did for the coin flips.

Let $F_k := \{ |S_1| < 3\alpha, \ldots, |S_{k-1}| < 3\alpha, |S_k| \ge 3\alpha \}$

Notice that the F_k's are disjoint, $\bigcup_{k=1}^{n} F_k = \{ \max_{1\le k\le n} |S_k| \geq 3\alpha \}$ and

$$\omega \in F_k \cap \{|S_n| < \alpha\} \Rightarrow |S_k(\omega)| \geq 3\alpha \ \& \ |S_n(\omega)| < \alpha$$

$$\overset{(*)}{\Rightarrow} |S_n(\omega) - S_k(\omega)| > 2\alpha$$

and $\omega \in F_k$

Now

$$P\left(\max_{1\le k\le n} |S_k| \geq 3\alpha\right)$$

$$= P\left(\max_{1\le k\le n} |S_k| \geq 3\alpha,\ |S_n| \geq \alpha\right) + P\left(\max_{1\le k\le n} |S_k| \geq 3\alpha,\ |S_n| < \alpha\right)$$

$$\leq P(|S_n| \geq \alpha) + \sum_{k=1}^{n-1} P(F_k \cap \{|S_n| < \alpha\})$$

drop $k=n$ term since $F_n \cap \{|S_n| < \alpha\} = \emptyset$

$$\leq P(|S_n| \geq \alpha) + \sum_{k=1}^{n-1} P(F_k \cap \{|S_n - S_k| > 2\alpha\})$$

by (*)

$$= P(|S_n| \geq \alpha) + \sum_{k=1}^{n-1} P(F_k) P(|S_n - S_k| > 2\alpha)$$

since $F_k \subset \sigma\langle X_1, \dots, X_k\rangle$ and $\{|S_n - S_k| > 2\alpha\} \subset \sigma\langle X_{k+1}, \dots, X_n\rangle$

$$\leq P(|S_n| \geq \alpha) + \max_{1\le k\le n} \underbrace{P(|S_n - S_k| > 2\alpha)}_{\leq P(|S_n| > \alpha) + P(|S_k| > \alpha)}$$

$$\leq 3 \max_{1\le k\le n} P(|S_k| \geq \alpha).$$

Q.E.D.

Characterizing and Constructing probability measures with densities

In undergraduate probability we are taught that probability densities characterize "continuous r.v.s" & probability mass functions characterize "discrete r.v.s".

e.g.

if $P(X \in B) = \int_B \underbrace{e^{-x} I_{(0,\infty)}(x)}_{\text{then this is the density}} dx$

but if

$$P(X \in B) = \sum_{\substack{k \in B \\ k = 0,1,\dots,n}} \underbrace{\binom{n}{k} \left(\tfrac{1}{3}\right)^k \left(\tfrac{2}{3}\right)^{n-k}}_{\text{then this is the probability mass function}}$$

It was annoying to me that you had to figure something out about X, i.e. continuous or discrete, before trying to compute probabilities. Our general integration theory unifies both cases and gives an accessible method for characterizing and constructing prob measures.

For the rest of this section let $(\Omega, \mathcal{F}, \mu)$ denote a measure space (unless stated otherwise).

Def: If $f \in Q(\Omega, \mathcal{F}, \mu)$ then the set function $\int_{\cdot} f\,d\mu : \mathcal{F} \to \overline{\mathbb{R}}$ mapping

$$A \in \mathcal{F} \mapsto \int_A f\,d\mu := \int_\Omega f I_A \,d\mu$$

is called the <u>indefinite integral of f with respect to μ</u>.

Theorem ($\int_{\cdot} f\,d\mu$ is σ-additive)

If $f \in Q(\Omega, \mathcal{F}, \mu)$ then $\int_{\cdot} f\,d\mu$ is countably additive over disjoint $\mathcal{F}$-sets.

Proof: Let $F_1, F_2, \ldots$ be disjoint $\mathcal{F}$-sets. Use the "2-3" argument.

Step "2": Suppose $f \in \mathcal{M}(\Omega, \mathcal{F})$.

$$\int_{\cup_k F_k} f\,d\mu = \int_\Omega \sum_{k=1}^\infty f I_{F_k}\,d\mu$$

since F_k's disjoint

$$= \int_\Omega \lim_n \uparrow \sum_{k=1}^n f I_{F_k}\,d\mu$$

since these are ≥ 0

$$\overset{\text{Big 3(3)}}{=} \lim_n \uparrow \int_\Omega \sum_{k=1}^n f I_{F_k}\,d\mu$$

$$\overset{\text{Big 3(3)}}{=} \lim_n \uparrow \sum_{k=1}^n \int_\Omega f I_{F_k}\,d\mu$$

$$= \sum_{k=1}^\infty \int_{F_k} f\,d\mu$$

$\therefore$ the Thm holds over $\mathcal{M}(\Omega, \mathcal{F})$.

Step "3": Suppose $f \in Q(\Omega, \mathcal{F}, \mu)$.

Certainly $f I_{\cup_k F_k} \in Q(\Omega, \mathcal{F}, \mu)$ since

$$(f I_{\cup_k F_k})^{\pm} = f^{\pm} I_{\cup_k F_k} \leq f^{\pm}$$

$$\therefore \int_{\cup_k F_k} f\,d\mu \overset{\text{def}}{=} \int_{\cup_k F_k} f^+\,d\mu - \int_{\cup_k F_k} f^-\,d\mu$$

$$\overset{\text{step "2"}}{=} \underbrace{\sum_{k=1}^\infty \int_{F_k} f^+\,d\mu}_{} - \underbrace{\sum_{k=1}^\infty \int_{F_k} f^-\,d\mu}_{}$$

one of these is finite. & Both are defined

$$= \sum_{k=1}^\infty \left[\int_{F_k} f^+\,d\mu - \int_{F_k} f^-\,d\mu\right]$$

by Big 3 for counting measure.

$$= \sum_{k=1}^\infty \int_{F_k} f\,d\mu.$$

Q.E.D.

Corollary:

i) If $f \in \mathcal{M}(\Omega, \mathcal{F})$ then

$\int_{\cdot} f\,d\mu$ is a measure on $(\Omega, \mathcal{F})$

ii) If $f \in \mathcal{M}(\Omega, \mathcal{F})$ & $\int_\Omega f\,d\mu = 1$ then

$\int_{\cdot} f\,d\mu$ is a probability measure on $(\Omega, \mathcal{F})$.

Definition:

For any two measures μ, ν on $(\Omega, \mathcal{F})$ if $\exists\, \delta \in \mathcal{M}(\Omega, \mathcal{F})$ s.t.

$$\nu(A) = \int_A \delta\,d\mu, \quad \forall A \in \mathcal{F}$$

Then <u>δ is a density of ν with respect to μ</u>.

Here is an important question:

$\boxed{\text{Are densities unique?}}$

the next example shows the answer must be

$\boxed{\text{Not without assumptions}}$

e.g. Let $\Omega = \mathbb{R}$
$\mathcal{F} = \{\emptyset, \Omega, (-\infty, 0), [0, \infty)\}$

$\therefore \mathcal{L}'(A) = \int_A 1\, d\mathcal{L}' = \int_A 2\, d\mathcal{L}' \quad \forall A \in \mathcal{F}$

these are both densities w.r.t $\mathcal{L}'$ resulting in the same measure.

The following theorem gives sufficient conditions for uniqueness (as a corollary)

Thm: Suppose $f, g \in Q(\Omega, \mathcal{F}, \mu)$. If $f \in L_1(\Omega, \mathcal{F}, \mu)$ or $g \in L_1(\Omega, \mathcal{F}, \mu)$ or μ is σ-finite then

$$\int_\bullet f\, d\mu \le \int_\bullet g\, d\mu \text{ on } \mathcal{F} \iff f \le g \quad \mu\text{-a.e.}$$

Proof:

$\Leftarrow$: Follows directly from Big 3 [1]

$\Rightarrow$:

Case 1: Suppose $f \in L_1(\Omega, \mathcal{F}, \mu)$ or $g \in L_1(\Omega, \mathcal{F}, \mu)$. We will show $\mu(f > g) = 0$ by the "indicate what you want trick"!

$f I_{\{f>g\}} \ge g I_{\{f>g\}}$

$\Rightarrow \int_{f>g} f\, d\mu \ge \int_{f>g} g\, d\mu$ by Big 3 [1].

$\Rightarrow \int_{f>g} f\, d\mu = \int_{f>g} g\, d\mu$ since $\int_\bullet f\, d\mu \le \int_\bullet g\, d\mu$ on $\mathcal{F}$.

$\Rightarrow \int \underbrace{(f-g) I_{\{f>g\}}}_{\ge 0}\, d\mu = 0$ by Big 3 [2] since $f \in L_1$ or $g \in L_1$

$\Rightarrow (f-g) I_{\{f>g\}} = 0$ μ-a.e. by a.e. useful facts

$\Rightarrow \mu(f > g) = 0$

since $f(\omega) > g(\omega) \Rightarrow (f(\omega) - g(\omega)) I_{\{f>g\}}(\omega) = 1$

Case 2: Suppose μ is finite.

First notice that

$f \le g$ on $\{f = \infty\} \cap \{g = \infty\}$ (1)

$f \le g$ on $\{f = -\infty\} \cap \{g = \infty\}$ (2)

$f \le g$ on $\{f = -\infty\} \cap \{g = -\infty\}$ (3)

We also have

$\mu(\{f = \infty\} \cap \{g = -\infty\}) = 0$ (4)

otherwise it would contradict the assumption $\int_\bullet f\, d\mu \le \int_\bullet g\, d\mu$. Now we just show

$f \le g$ μ-a.e. on $\{|f| < \infty\} \cup \{|g| < \infty\}$. (5)

Let $A_n := \{|f| < n\}$. Since μ is a finite measure

$f I_{A_n} \in L_1$ & $g I_{A_n} \in Q$.

Also

$$\int_\bullet f I_{A_n}\, d\mu = \int_{\bullet \cap A_n} f\, d\mu \le \int_{\bullet \cap A_n} g\, d\mu = \int_\bullet g I_{A_n}\, d\mu$$

on $\mathcal{F}$. Since $f I_{A_n} \in L_1$ case 1 implies

$f I_{A_n} \le g I_{A_n}$ μ-a.e.

i.e. $f \le g$ μ-a.e. on $\{|f| < n\}$

$\therefore f \le g$ μ-a.e. on $\{|f| < \infty\} = \bigcup_{n=1}^\infty \{|f| < n\}$.

A similar argument shows

$f \le g$ μ-a.e. on $\{|g| < \infty\}$.

Now the union of (1)-(5) gives

$f \le g$ μ-a.e.

Case 3: Suppose μ is σ-finite.

Let $F_k \in \mathcal{F}$ s.t. $\mu(F_k) < \infty$ and $\bigcup_{k=1}^\infty F_k = \Omega$.

$\therefore \mu(f > g) = \sum_{k=1}^\infty \underbrace{\mu(\{f > g\} \cap F_k)}_{= \mu_k(f > g)}$ where

$\mu_k(\bullet) := \mu(\bullet \cap F_k)$

Now case 2 applies to the finite measure μ_k since

$$\int_\bullet f\,d\mu_k = \int_\bullet f I_{F_k}\,d\mu \quad \text{by a "1-2-3" argument}$$

$$= \int_{\bullet \cap F_k} f\,d\mu$$

$$\leq \int_{\bullet \cap F_k} g\,d\mu = \int_\bullet g\,d\mu_k$$

$\therefore$ case 2 implies $\mu_k(f>g)=0$.

$\therefore$ $\mu(f>g)=0$.

QED

Corollary (uniqness of densities)

Let $f,g \in \mathcal{Q}(\Omega,\mathcal{F},\mu)$. If f or g is integrable or μ is σ-finite then

$$\int_\bullet f\,d\mu = \int_\bullet g\,d\mu \text{ on } \mathcal{F} \iff f=g \ \mu\text{-a.e.}$$

Note: If P is a probability measure and μ is a measure (over $(\Omega,\mathcal{F})$) then

i) $P(\bullet) = \int_\bullet \delta\,d\mu \Rightarrow$ both $\int_\Omega \delta^+ d\mu < \infty$ and $\int_\Omega \delta^- d\mu < \infty$.

$\Rightarrow \delta \in L_1(\Omega,\mathcal{F},\mu)$

$\Rightarrow \delta$ is unique μ-a.e.

ii) $\mu(\bullet) = \int_\bullet \delta\,dP \Rightarrow \delta$ is unique P-a.e. since P is σ-finite.

The next theorem shows how to compute $\int_\Omega f\,d\nu$ when ν has density δ w.r.t. μ.

Theorem (slap in the density: $d\nu = \delta\,d\mu$)

Let ν and μ be densities on $(\Omega,\mathcal{F})$ where ν has density δ w.r.t μ.

Then

$$f \in \mathcal{Q}^\pm(\Omega,\mathcal{F},\nu) \iff f\delta \in \mathcal{Q}^\pm(\Omega,\mathcal{F},\mu)$$

and either one implies

$$\int_\Omega f\,d\nu = \int_\Omega f\delta\,d\mu$$

i.e. $d\nu = \delta\,d\mu$
i.e. $\delta = \frac{d\nu}{d\mu}$

Proof: Again use "1-2-3 argument".

Step 1: Suppose $f \in \mathcal{M}_s(\Omega,\mathcal{F})$ so that $f = \sum_{k=1}^n c_k I_{A_k}$ for $A_k \in \mathcal{F}$ & $c_k \geq 0$.

Since $f \in \mathcal{M}(\Omega,\mathcal{F})$ clearly

$$f \in \mathcal{Q}^-(\Omega,\mathcal{F},\nu) \iff f\delta \in \mathcal{Q}^-(\Omega,\mathcal{F},\mu).$$

$$\therefore \int_\Omega f\,d\nu = \sum_{k=1}^n c_k \nu(A_k)$$

$$= \sum_{k=1}^n c_k \int_{A_k} \delta\,d\mu \quad \text{since } \delta \text{ is a density for } \nu \text{ w.r.t. } \mu$$

$$= \int_\Omega \underbrace{\left(\sum_{k=1}^n c_k I_{A_k}\right)}_{f} \delta\,d\mu \quad \text{by little 3}$$

and this implies

$$f \in \mathcal{Q}^+(\Omega,\mathcal{F},\nu) \iff f\delta \in \mathcal{Q}^+(\Omega,\mathcal{F},\mu).$$

Step 2: Suppose $f \in \mathcal{M}(\Omega,\mathcal{F})$. Then the result follows similarly by little 3.

Step 3: From step 2

$$\int_\Omega f^\pm\,d\nu = \int_\Omega f^\pm \delta\,d\mu = \int_\Omega (f\delta)^\pm d\mu$$

$\therefore f \in \mathcal{Q}^\pm(\Omega,\mathcal{F},\nu) \iff f\delta \in \mathcal{Q}^\pm(\Omega,\mathcal{F},\mu)$ QED.

Notation:

Suppose ν and μ are measures on $(\Omega, \mathcal{F})$. If ν has a density w.r.t μ I will denote it:

$\frac{d\nu}{d\mu}$ ← Non-negative ⓜ function mapping $\Omega \to \bar{\mathbb{R}}$ s.t. $\nu(\cdot) = \int_{\cdot} \frac{d\nu}{d\mu} d\mu$.

Moreover when I say $\frac{d\nu}{d\mu}$ exists I mean there exists some density of ν w.r.t. μ (it will be unique μ-a.e. if $\frac{d\nu}{d\mu} \in L_1(\Omega, \mathcal{F}, \mu)$ or μ is σ-finite.

Theorem (chain rule)

Suppose ν, ρ, μ are measures on $(\Omega, \mathcal{F})$ with μ σ-finite. If $\frac{d\rho}{d\nu}$ and $\frac{d\nu}{d\mu}$ exists then

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

↳ even if μ isn't σ-finite this serves as a density of ρ w.r.t. μ

Proof:

$$\int_{\cdot} \frac{d\rho}{d\nu}\frac{d\nu}{d\mu} d\mu = \int_{\cdot} \frac{d\rho}{d\nu} d\nu \quad \text{by slap in the density}$$

$\in \mathcal{M}(\Omega, \mathcal{F})$

$$= \int_{\cdot} d\rho \quad \text{by slap in the density}$$

$$= \rho(\cdot) \quad \text{def}$$

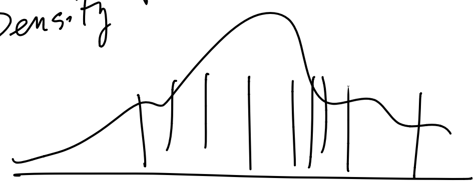
over $\mathcal{F}$.

Uniqueness follows since μ is σ-finite.

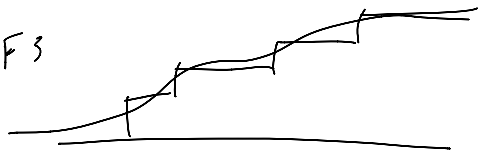
QED

Glavenko-Cantelli

Density vrs empirical measure



CDFs



Theorem (Glavenko-Cantelli)

- wilks formula.
- densities, sheffe ... positive def. application
- poisson likelihood.