Expectations and Independence

Theorem Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $X_1, \dots, X_n : \Omega \to \mathbb{R}$ be random variables.
Then $X_1, \dots, X_n$ are independent $\iff$ for any Borel functions $f_1, \dots, f_n : \mathbb{R} \to \mathbb{R}$, $\mathbb{E}\left[\prod_{k=1}^{n} f_k(X_k)\right] = \prod_{i=1}^{n} \mathbb{E}[f_k(X_k)]$

Recall lemma: **Lemma** Let $f : \Omega \to [0, \infty)$. Then there exist simple f^n's $(f_n\ n \geq 1)$ s.t. $0 \leq f_n \uparrow f$, and s.t. f_n is $\sigma(f)/\mathcal{B}(\mathbb{R})$-measurable for all $n \geq 1$.

Proof. Set $f_n = \sum_{k=0}^{n \cdot 2^n - 1} \frac{k}{2^n} \cdot \mathbb{1}_{\{\frac{k}{2^n} \leq f < \frac{k+1}{2^n}\}}$. $\square$

Proof of Theorem First suppose that for any Borel functions $f_1, \dots, f_n : \mathbb{R} \to \mathbb{R}$, $\mathbb{E}\left[\prod_{k=1}^{n} f_k(X_k)\right] = \prod_{i=1}^{n} \mathbb{E}[f_k(X_k)]$

Now fix sets $(E_k \in \sigma(X_k),\ 1 \leq k \leq n)$.

Since $\sigma(X_k) = \{X_k^{-1}(B) : B \in \mathcal{B}(\mathbb{R})\}$, we can find sets $B_1, \dots, B_n \in \mathcal{B}(\mathbb{R})$ such that $E_k = \{X_k \in B_k\}$ for $1 \leq k \leq n$.

Then taking $f_k = \mathbb{1}_{B_k}$, we have

$$\prod_{k=1}^{n} \mathbb{E}[f_k(X_k)] = \prod_{k=1}^{n} \mathbb{E}[\mathbb{1}_{B_k}(X_k)] = \prod_{k=1}^{n} \mathbb{P}(X_k \in B_k) = \prod_{k=1}^{n} \mathbb{P}(E_k)$$

$$\| $$

$$\mathbb{E}\left[\prod_{k=1}^{n} f_k(X_k)\right] = \mathbb{E}\left[\prod_{k=1}^{n} \mathbb{1}_{B_k}(X_k)\right] = \mathbb{P}\left(\bigcap_{k=1}^{n} \{X_k \in B_k\}\right) = \mathbb{P}\left(\bigcap_{k=1}^{n} E_k\right).$$ Thus $X_1, \dots, X_n$ are independent.

Next suppose that $X_1, \dots, X_n$ are independent. Then for all $B_1, \dots, B_n \in \mathcal{B}_n$, $\mathbb{E}\left[\prod_{k=1}^{n} \mathbb{1}_{B_k}(X_k)\right] = \prod_{k=1}^{n} \mathbb{E}[\mathbb{1}_{X_k}(B_k)]$

Now let $f_1, \dots, f_n$ be simple Borel functions; we may write $f_k = \sum_{\ell=1}^{m} c_{k,\ell} \mathbb{1}_{B_{k,\ell}}$, for some $c_{k,\ell} \in \mathbb{R}$ and $B_{k,\ell} \in \mathcal{B}(\mathbb{R})$.

Then

$$\mathbb{E}\Big[\prod_{k=1}^{n} f_k(X_k)\Big] = \mathbb{E}\Big[\prod_{k=1}^{n}\Big(\sum_{\ell=1}^{m} c_{k,\ell}\,\mathbb{1}_{B_{k,\ell}}(X_{k,\ell})\Big)\Big] \overset{\text{L. of E.}}{=} \sum_{\ell_1,\dots,\ell_n=1}^{m} \mathbb{E}\prod_{k=1}^{n} c_{k,\ell_k}\,\mathbb{1}_{B_{k,\ell_k}}(X_k)$$

$$= \sum_{\ell_1,\dots,\ell_n=1}^{m} \prod_{k=1}^{n} c_{k,\ell_k}\,\mathbb{E}\,\mathbb{1}_{B_{k,\ell_k}} = \prod_{k=1}^{n} \mathbb{E}\sum_{\ell=1}^{m} c_{k,\ell}\,\mathbb{1}_{B_{k,\ell}} = \prod_{k=1}^{n}\mathbb{E} f_k(X_k)$$

So this gives the factorization formula for simple functions.

Next suppose $f_1,\dots,f_n$ are non-negative Borel functions, and set $Y_k = f_k(X_k)$. Then $Y_1,\dots,Y_n$ are independent. By the lemma, we may find simple functions $(Y_{n,m}, m\geq 1)$ s.t. $0\leq Y_{n,m}\nearrow Y_n$ with $Y_{n,m}$ $\sigma(Y_n)/\mathcal{B}(\mathbb{R})$-measurable.

By the factorization formula for simple functions, and the monotone convergence theorem,

$$\mathbb{E}\prod_{k=1}^{n} f_k(X_k) = \mathbb{E}\prod_{k=1}^{n} Y_k \overset{\text{Mon. conv. thm}}{=} \lim_{m\to\infty}\mathbb{E}\prod_{k=1}^{n} Y_{k,m} \overset{\text{F.F.}}{=} \lim_{m\to\infty}\prod_{k=1}^{n}\mathbb{E}Y_{k,m} \overset{\text{Mon. conv. thm.}}{=} \prod_{k=1}^{n}\mathbb{E}Y_k = \prod_{k=1}^{n}\mathbb{E} f_k(X_k).$$

This establishes the factorization formula for non-negative functions.

Finally, if $f_1,\dots,f_n$ are bounded and measurable then we can again use linearity to write $f_k(X_k) = Y_k = Y_k^+ - Y_k^-$, and we then have

$$\mathbb{E}\prod_{k=1}^{n} f_k(X_k) = \mathbb{E}\prod_{k=1}^{n}(Y_k^+ - Y_k^-) = \sum_{(\sigma_1,\dots,\sigma_n)\in\{+,-\}^n} (-1)^{\#\{i:\sigma_i=-\}}\,\mathbb{E}\Big[\prod_{k=1}^{n} Y_k^{\sigma_k}\Big]$$

$$\overset{\text{non-neg. case}}{=} \sum_{(\sigma_1,\dots,\sigma_n)\in\{+,-\}^n} (-1)^{\#\{i:\sigma_i=-\}} \prod_{k=1}^{n}\mathbb{E}\big[Y_k^{\sigma_k}\big] = \prod_{k=1}^{n}\mathbb{E}\big[Y_k^+ - Y_k^-\big] = \prod_{k=1}^{n}\mathbb{E}\big[f_k(X_k)\big]. \quad \square$$

$\boxed{\text{Corollary}}$ If $X, Y : \Omega \to \mathbb{R}$ are independent and $\underbrace{X \in L^1(\mathbb{P}),\ Y \in L^1(\mathbb{P})}_{\mathbb{E}|X| < \infty,\ \mathbb{E}|Y| < \infty}$,
then $XY \in L^1(\mathbb{P})$ and $\mathbb{E}XY = \mathbb{E}X\,\mathbb{E}Y$.

$\boxed{\text{Proof}}$ Notation: for a function $f : \Omega \to \mathbb{R}$ write $f_{\leq n} := f \mathbb{1}_{|f| \leq n}$.

First suppose $X, Y \geq 0$. Then by Mon. conv. thm and independence,

$$\mathbb{E}[XY] = \lim_{n\to\infty} \mathbb{E}[X_{\leq n} Y_{\leq n}] = \lim_{n\to\infty} \mathbb{E}X_{\leq n}\,\mathbb{E}Y_{\leq n} = \mathbb{E}X\,\mathbb{E}Y.$$

Write $X = X^+ - X^-$, $Y = Y^+ - Y^-$;

Then $\mathbb{E}|XY| = \mathbb{E}X^+Y^+ + \mathbb{E}X^+Y^- + \mathbb{E}X^-Y^+ + \mathbb{E}X^-Y^-$

non-neg. case $\;= (\mathbb{E}(X^+ + X^-))(\mathbb{E}(Y^+ + Y^-)) = \mathbb{E}|X|\,\mathbb{E}|Y| < \infty$,

So $XY \in L^1(\mathbb{P})$ and

$$\mathbb{E}XY = \mathbb{E}X^+Y^+ \quad \mathbb{E}X^+Y^- \quad \mathbb{E}X^-Y^+ + \mathbb{E}Y^+Y^-$$

$$= (\mathbb{E}(X^+ - X^-))(\mathbb{E}(Y^+ - Y^-)) = \mathbb{E}X\,\mathbb{E}Y. \quad \square$$

Monotone class theorem Fix a measurable space $(\Omega, \mathcal{F})$ let $P \subset \mathcal{F}$ be a π-system over Ω, with $\Omega \in P$. Let $\mathcal{A}$ be a collection of functions $f: \Omega \to \mathbb{R}$ s.t. the following all hold.

(a) $\mathbb{1}_P \in \mathcal{A}$ for all $P \in P$;

(b) If $f, g \in \mathcal{A}$ and $c \in \mathbb{R}$ then $cf + g \in \mathcal{A}$

(c) If $(f_n\ n \geq 1)$ are in $\mathcal{A}$ and $0 \leq f_n \uparrow f$ and f is bounded then $f \in \mathcal{A}$.

Then $\mathcal{A} \supseteq \{f: \Omega \to \mathbb{R} : f \text{ is } (\sigma(P)/\mathcal{B}(\mathbb{R}))\text{-measurable}\}$.

Proof Step 1 Let $\Lambda := \{F \subset \Omega : \mathbb{1}_F \in \mathcal{A}\}$. Then $P \subset \Lambda$ by definition.

Also, if $E, F \in \Lambda$ and $E \subset F$ then $\mathbb{1}_E, \mathbb{1}_F \in \mathcal{A}$ so $\overbrace{\mathbb{1}_{F \setminus E} = \mathbb{1}_F - \mathbb{1}_E}^{(b)} \in \mathcal{A}$ so $F \setminus E \in \Lambda$.

Moreover, if $0 \leq E_n \uparrow E$ and each $E_n \in \Lambda$ then $\mathbb{1}_{E_n} \uparrow \mathbb{1}_E$ so $\mathbb{1}_E \in \mathcal{A}$ so $E \in \Lambda$.

Thus Λ is a λ-system so $\sigma(P) \subseteq \Lambda$ by Dynkin's π-system lemma.

Step 2
- By Step 1, For all $E \in \sigma(P)$, $\mathbb{1}_E \in \mathcal{A}$.
- By (b), for any $E_1, \dots, E_n \in \sigma(P)$ and $c_1, \dots, c_n \in \mathbb{R}$, $\sum_{i=1}^{n} c_i \mathbb{1}_{E_i} \in \mathcal{A}$, so $\mathcal{A}$ contains all simple $(\sigma(P)/\mathcal{B}(\mathbb{R}))$-measurable functions.
- If $f: \Omega \to [0, \infty)$ is $(\sigma(P)/\mathcal{B}(\mathbb{R}))$-measurable then f is an increasing limit of simple f^ns. $0 \leq f_n \uparrow f$, so if f is also bounded then $f \in \mathcal{A}$ by (c). So $\mathcal{A} \supseteq \{\text{Non-neg. bdd. } \sigma(P)/\mathcal{B}(\mathbb{R})\text{-meas. } f\}$
- Any $f: \Omega \to \mathbb{R}$ can be written as $f = f^+ - f^-$ and if f is $\sigma(P)$-$\mathcal{B}(\mathbb{R})$ meas. and bounded then f^+, f^- are $(\sigma(P)/\mathcal{B}(\mathbb{R}))$-meas., non-neg. and bdd so $f^+, f^- \in \mathcal{A}$ so $f \in \mathcal{A}$ by (b). $\square$

The probabilistic method

To prove an object with a given property P exists, one may construct a random variable X with the property that $\mathbb{P}$(X has property P) > 0.

Examples

① Existence of continuous nowhere differentiable functions

Let $D_n = \{\frac{i}{2^n}, 0 \le i \le 2^n\}$, $D_{\le n} = \bigcup_{i \le n} D_i$, $D = D_\infty$. Let $(N_x, x \in D)$ be IID $N(0,1)$ r.v.s.

Define $B_1(0) = 0$ $\quad B_1(1) = N_1$ $\quad B_1(x) = x N_1$ (linear interpolation)

Inductively, given B_{n-1}, let

$$B_n(x) = \begin{cases} B_{n-1}(x) & \text{if } x \in D_{\le n-1} \\ B_{n-1}(x) + N_x/\sqrt{2^n} & \text{if } x \in D_n \setminus D_{\le n-1} \\ \text{linear interpolation between points of } D_{\le n}. \end{cases}$$

Then it turns out that $\mathbb{P}(B_n$ is a uniformly convergent function$) = 1$.

The limit B_∞ is Brownian motion on $[0,1]$

And $\mathbb{P}(B_\infty$ is nowhere differentiable$) = 1$.

② Small-discrepancy signings of vectors

Prop Let $v_1, \dots, v_n \in \mathbb{R}^n$ have $|v_i| = 1$. Then there exist $\sigma_1, \dots, \sigma_n \in \{-1, 1\}$ s.t. $|\sigma_1 v_1 + \dots + \sigma_n v_n| \leq \sqrt{n}$.

Proof

Let $\sigma_1, \dots, \sigma_n$ be independent and uniform on $\{-1,1\}$. Set $X = |\sigma_1 v_1 + \dots + \sigma_n v_n|^2$.

Then $X = (\sigma_1 v_1 + \dots + \sigma_n v_n) \cdot (\sigma_1 v_1 + \dots + \sigma_n v_n)$

$$\mathbb{E}[X] = \mathbb{E} \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_i \sigma_j v_i \cdot v_j = \sum_{i=1}^{n} \sum_{j=1}^{n} v_i \cdot v_j \cdot \mathbb{E}[\sigma_i \sigma_j]$$

$$= \sum_{i=1}^{n} |v_i|^2 \cdot \mathbb{E}[\sigma_i^2] + 2 \sum_{1 \leq i < j \leq n} v_i \cdot v_j \, \mathbb{E}[\sigma_i \sigma_j]$$

$$= \sum_{i=1}^{n} 1 + 2 \sum_{1 \leq i < j \leq n} 0 = n.$$

So $\mathbb{E} X = n$ so $\mathbb{P}(|\sigma_1 v_1 + \dots + \sigma_n v_n| \leq \sqrt{n}) = \mathbb{P}(X \leq n) > 0$

So there must be at least one choice of $\sigma_1, \dots, \sigma_n$ s.t. $|\sigma_1 v_1 + \dots + \sigma_n v_n| \leq \sqrt{n}$.

... but how do we actually do stuff?

$f_{\le n} := f \mathbb{1}_{|f| \le n}$

★ If X is a r.v. which takes values in $\mathbb{N}$ (or some other countable set $N \subset [0,\infty)$)

then $\mathbb{E}X = \lim_{n\to\infty} \mathbb{E}X_{\le n} = \lim_{n\to\infty} \int X_{\le n}\, d\mathbb{P}$; $X_{\le n}$ is simple so

$$\mathbb{E}X_{\le n} = \sum_{k=0}^{n} k \cdot \mathbb{P}(X=k) = \sum_{k=0}^{n} k \cdot \mu_X(k)$$ ← by def. of μ_X and so $\mathbb{E}X = \sum_{k\in\mathbb{N}} k \cdot \mathbb{P}(X=k)$.

So, for example, if $P \sim \text{Poisson}(\lambda)$ then $\mathbb{E}P = \sum_{k\in\mathbb{N}} k \cdot \frac{\lambda^k e^{-\lambda}}{k!} = \lambda \cdot \sum_{k\ge 1} \frac{\lambda^{k-1} e^{-\lambda}}{(k-1)!} = \lambda$

★ Suppose I tell you $N = (N_i, i \ge 1)$ are IID $N(0,1)$ and $P \sim \text{Poisson}(1)$ is independent of N.

How do you compute $\mathbb{P}\left(\sum_{i=1}^{P} N_i \ge 1\right) = \mathbb{E}\left[\mathbb{1}_{\sum_{i=1}^{P} N_i \ge 1}\right] = \int \mathbb{1}_{\sum_{i=1}^{P} N_i}(\omega)\, \mathbb{P}(d\omega)$?

Using <u>densities</u>, <u>change of variables</u>, and <u>Fubini's Theorem</u>.

Definitions Given a measure space $(\Omega, \mathcal{F}, \mu)$ and $f: \Omega \to \mathbb{R}$ non-negative and measurable, define a new measure μf on $(\Omega, \mathcal{F})$ by $\mu f(A) = \int_A f\, d\mu := \int f \mathbb{1}_A\, d\mu$.

$\boxed{\text{Examples}}$

- Let X be a non-negative r.v. on $(\Omega, \mathcal{F}, \mathbb{P})$, for $\lambda \ge 0$ let $f_\lambda(x) = e^{-\lambda x}$.

Then $\mu f_\lambda(X)(A) = \int_A e^{-\lambda X}\, d\mathbb{P} = \mathbb{E}[e^{-\lambda X} \mathbb{1}_A]$

The function $\lambda \longmapsto \mu f_\lambda(X)(\Omega) = \mathbb{E}[e^{-\lambda X}]$ is the Laplace transform of X

Size-biasing

- Let X be a r.v. with $X \geq 0$, $\mathbb{E}X < \infty$. Recall μ_X is dist. of X.

Then $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_X)$ is a prob. space. The size-biased distribution of X is $\hat{\mu}_X := \left(\mu_X \frac{X}{\mathbb{E}X}\right)$, so $\hat{\mu}_X(A) = \left(\mu_X \frac{X}{\mathbb{E}X}\right)(A) = \mathbb{E}\left[\frac{X}{\mathbb{E}X} \mathbb{1}_{X \in A}\right]$

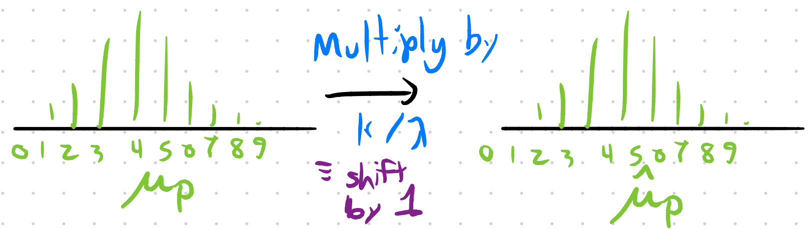
Note $\hat{\mu}_X(\mathbb{R}) = \mathbb{E}\left[\frac{X}{\mathbb{E}X}\right] = \frac{1}{\mathbb{E}X}\mathbb{E}[X] = 1$, so $\hat{\mu}_X$ is another prob. dist

- Let $P : \Omega \to \mathbb{R}$ be Poisson(λ); $\mathbb{P}(P = k) = \lambda^k e^{-\lambda}/k!$, $\mathbb{E}P = \lambda$.

Then $\hat{\mu}_P(A) = \left(\mu_P \frac{P}{\lambda}\right)(A) = \mathbb{E}\left[P/\lambda \; \mathbb{1}_{P \in A}\right]$

$$= \sum_{k \in A \cap \{1,2,3,\ldots\}} \frac{k}{\lambda} \cdot \mathbb{P}(P = k)$$

$$= \sum_{k \in A \cap \{1,2,3,\ldots\}} \frac{k}{\lambda} \cdot \frac{\lambda^k e^{-\lambda}}{k!} = \sum_{k \in A \cap \{1,2,3,\ldots\}} \frac{\lambda^{k-1} e^{-\lambda}}{(k-1)!} = \mathbb{P}(P + 1 \in A) = \mu_{P+1}(A)$$



"A size-biased Poisson(λ) dist. is just $1 +$ Poisson(λ)"

Definition If $\nu = \mu f$ then we say ν has density f with respect to μ.

Say X has density f with respect to Lebesgue measure if μ_X has density f w.r.t. Lebesgue, i.e. if $\forall B \in \mathcal{B}(\mathbb{R})$, $\mu_X(B) = \int \mathbb{1}_B \, \mu_X(dx) = \int_B f(x)\,dx$

The definition is justified by the following proposition.

Proposition If ν has density f with respect to μ then $\int g \, d\nu = \int g f \, d\mu$ holds for measurable g whenever a) $g \geq 0$ or b) $g \in L_1(\mu)$. Moreover, $g \in L^1(\nu) \iff gf \in L^1(\mu)$.

Proof If $g = \mathbb{1}_A$ for some $A \in \mathcal{F}$ then holds by definition. Then build up using monotone class thm and monotone convergence thm. $\square$

Change of variables formula

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a p.s. and let $X : \Omega \to \mathbb{R}$ be a r.v..

Then for all measurable $g : \mathbb{R} \to \mathbb{R}$ with $\mathbb{E}|g(X)| < \infty$, have $\mathbb{E}\, g(X) = \int_{\mathbb{R}} g(x)\, \mu(dx)$.

Moreover, if X has density f w.r.t. Lebesgue measure then $\mathbb{E}\, g(X) = \int_{\mathbb{R}} g(x) f(x)\, dx$.

Proof True for $\mathbb{1}_B(X)$, then build up using monotone class thm and monotone convergence thm. $\square$

Example If $N \sim N(0,1)$ then N has density $\Phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ w.r.t. Lebesgue measure

$|N|$ has density $\varphi(x) = \sqrt{\frac{2}{\pi}}\, e^{-x^2/2}\, \mathbb{1}_{x \geq 0}$ w.r.t. Lebesgue measure.

What about the size-biasing of $|N|$?

The distribution is $\hat{\mu}_{|N|} = \left(\mu_{|N|} \cdot \frac{|N|}{\mathbb{E}|N|}\right)$

$$\mathbb{E}|N| = \int |N|\, d\mathbb{P} \underset{\text{C of V}}{=} \int |x|\, d\mu_{|N|} \underset{\text{C of V}}{=} \int_{[0,\infty)} x \cdot \sqrt{\tfrac{2}{\pi}}\, e^{-x^2/2} dx = \left[-\sqrt{\tfrac{2}{\pi}}\, e^{-x^2/2}\right]_0^\infty = \sqrt{\tfrac{2}{\pi}}.$$

So for B Borel, $B \subset [0,\infty)$,

$$\hat{\mu}_{|N|}(B) = \left(\mu_{|N|} \cdot \frac{|N|}{\sqrt{2/\pi}}\right)(B) = \int \mathbb{1}_B \, d\left(\mu_N \cdot \frac{|N|}{\mathbb{E}|N|}\right) = \int \mathbb{1}_B \cdot \frac{N}{\mathbb{E}N}\, d\mu$$

$$= \int_B \frac{x}{\sqrt{2/\pi}} \sqrt{\tfrac{2}{\pi}}\, e^{-x^2/2} dx = \int_B x e^{-x^2/2} dx$$

So $\hat{\mu}_{|N|}$ has density $x e^{-x^2/2}\, \mathbb{1}_{x \geq 0}$ w.r.t. Lebesgue measure.

This is the Rayleigh distribution.

Exercise If X, Y are independent standard Normal then $\sqrt{X^2+Y^2}$ is Rayleigh distributed.

More generally, if $(\Omega, \mathcal{F}, \mu)$ measure space and $(S, \mathcal{A})$ is a measurable space, then for a $f: \Omega \to S$ an $(\mathcal{F}/\mathcal{A})$-measurable function we may define $\nu(E) = \mu(f^{-1}(E))$ for $E \in \mathcal{A}$.

Then for all non-negative $(\mathcal{A}/\mathcal{B}(\mathbb{R}))$-measurable functions $g: S \to \mathbb{R}$, $\int g \, d\nu = \int g \circ f \, d\mu$.

Special case: $\mathbb{E}\, g(X) = \int g(x) \, \mu(dx)$.

Another special case: if $X = (X_1, \dots, X_n)$ are real r.v.s defined on a common space and $g: \mathbb{R}^n \to [0, \infty)$ is measurable then

$$\mathbb{E}\, g(X) = \int_{\mathbb{R}^n} g(\vec{x}) \, \mu_X(d\vec{x})$$

$\leftarrow$ Makes sense since $X: \Omega \to \mathbb{R}^n$ is measurable so μ_X is defined and is a measure on $\mathbb{R}^n$.

Exercise Prove that if $\varphi: [a,b] \to \mathbb{R}$ is C_1 (continuously differentiable) and strictly $\uparrow$ then for all Borel functions $g: [\varphi(a), \varphi(b)] \to [0, \infty)$,

$$\int_{\varphi(a)}^{\varphi(b)} g(y) \, dy = \int_a^b g(\varphi(y)) \varphi'(y) \, dy.$$

(Note: If $g = \mathbb{1}_{[c,d]}$ for $[c,d] \subset [a,b]$ this says $\varphi(d) - \varphi(c) = \int_c^d \varphi'(y) \, dy$, which is the fundamental thm of calc. Then use monotone class theorem.)