Theorem. Change of variables formula. Let X be a random element of (S, \mathcal{S}) with distribution μ , i.e., $\mu(A) = P(X \in A)$. If f is a measurable function from (S, \mathcal{S}) to $(\mathbf{R}, \mathcal{R})$ so that $f \geq 0$ or $E|f(X)| < \infty$, then

$$Ef(X) = \int_{S} f(y)\mu(dy)$$

Remark. To explain the name, write h for X and $P \circ h^{-1}$ for μ to get

$$\int_{\Omega} f(h(\omega))dP = \int_{S} f(y)d(P \circ h^{-1})$$

Proof. We will prove this result by verifying it in four increasingly more general special cases that parallel the way that the integral was defined in Section 1.4. The reader should note the method employed, since it will be used several times below.

CASE 1: INDICATOR FUNCTIONS. If $B \in \mathcal{S}$ and $f = 1_B$ then recalling the relevant definitions shows

$$E1_B(X) = P(X \in B) = \mu(B) = \int_S 1_B(y)\mu(dy)$$

CASE 2: SIMPLE FUNCTIONS. Let $f(x) = \sum_{m=1}^{n} c_m 1_{B_m}$ where $c_m \in \mathbf{R}$, $B_m \in \mathcal{S}$. The linearity of expected value, the result of Case 1, and the linearity of integration imply

$$Ef(x) = \sum_{m=1}^{n} c_m E1_{B_m}(X)$$

$$= \sum_{m=1}^{n} c_m \int_{S} 1_{B_m}(y) \mu(dy) = \int_{S} f(y) \mu(dy)$$

Case 3: Nonegative Functions. Now if $f \ge 0$ and we let

$$f_n(x) = ([2^n f(x)]/2^n) \wedge n$$

where [x] = the largest integer $\leq x$ and $a \wedge b = \min\{a, b\}$, then the f_n are simple and $f_n \uparrow f$, so using the result for simple functions and the monotone convergence

theorem:

$$Ef(X) = \lim_{n} Ef_n(X) = \lim_{n} \int_{S} f_n(y)\mu(dy) = \int_{S} f(y)\mu(dy)$$

CASE 4: INTEGRABLE FUNCTIONS. The general case now follows by writing $f(x) = f(x)^+ - f(x)^-$. The condition $E|f(X)| < \infty$ guarantees that $Ef(X)^+$ and $Ef(X)^-$ are finite. So using the result for nonnegative functions and linearity of expected value and integration:

$$Ef(X) = Ef(X)^{+} - Ef(X)^{-} = \int_{S} f(y)^{+} \mu(dy) - \int_{S} f(y)^{-} \mu(dy)$$
$$= \int_{S} f(y) \mu(dy)$$

which competes the proof.