Continuous Random Variables

Recall that a discrete random variable $X : S \to \mathbb{R}$ has either finite or at most countable range $R(X) = \{x_i \mid i = 1, \dots, n\}$

or

$= \{x_i \mid i \in \mathbb{N}\}$

and in this case we are interested in probabilities of the type

$$P(X = a) \stackrel{\text{def}}{=} P(X^{-1}(a))$$

or

$$P(X \leq a) = P(X^{-1}(-\infty, a])$$

etc., in general

$$\boxed{P(X \in B) \stackrel{\text{def}}{=} P(X^{-1}(B))}$$

where $B \subseteq R(X)$. Since $B \subseteq R(X)$ is at most countable, we can write

$$B = \{ b_i \in \mathbb{R} \mid X(s) = b_i \in R(X) \text{ for some } i \in \mathbb{N} \}$$

and therefore

$$X^{-1}(B) = X^{-1}\Big(\bigsqcup_{b_i \in B} b_i\Big)$$

$$= \bigsqcup_{b_i \in B} X^{-1}(b_i)$$

and therefore

$$P(X \in B) = P(X^{-1}(B))$$

$$= P\Big(\bigsqcup_{b_i \in B} X^{-1}(b_i)\Big)$$

a finite sum
OR
an infinite series

$$= \sum_{b_i \in B} P(X^{-1}(b_i))$$

$$= \sum_{b_i \in B} P(X = b_i)$$

$$= \sum_{b_i \in B} p(b_i)$$

pmf

In the continuous case, $R(X)$ is uncountable, so typically $B \subseteq R(X)$ will also be uncountable.

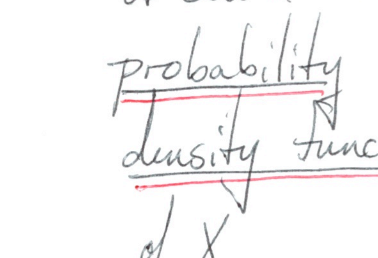
Therefore the sum $\sum_{b_i \in B} p(b_i) \equiv \sum_{b_i \in B} P(X = b_i)$ will have to be replaced by an integral

$f: \underset{\subseteq \mathbb{R}}{R(X)} \to \mathbb{R}$ is called the probability density function of X.

$$\boxed{\begin{aligned} P(X \in B) &= \int_B f(x)\,dx \\ &= \lim_{n \to \infty} \sum_{i=1}^{n} f(x_i^*)\,\Delta x \\ &\approx \underbrace{\sum_{i=1}^{n} f(x_i^*)\,\Delta x}_{\text{similar to the finite sum } \sum_{b_i \in B} p(b_i) \text{ in the discrete case}} \end{aligned}}$$

$\left(x_i^* \,\&\, x_i \in B \right)$

Alternatively, we can view the discrete sum $\sum_{b_i \in B} p(b_i)$ as an integral,



$$\sum_{b_i \in B} p(b_i) = \sum_{b_i \in B} \underbrace{p(b_i) \cdot \overset{\text{like } \Delta x}{1}}_{= \int_k^{k+1} p(b_i)\,dx} = \int_B p(x)\,dx$$

The difference between these two points of view is precisely the difference between the Riemann integral (used in undergrad prob.) and the Lebesgue integral (used in grad prob.)

Again, if X is discrete, we have

$$E[X] = \sum_{x \in R(X)} x\, p(x)$$

$$\equiv \sum_{x \in R(X)} x\, P(X = x)$$

&

$$E[g(X)] = \sum_{x \in R(X)} g(x)\, p(x)$$

In the continuous case, therefore, we have

$$\boxed{E[X] \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} x\, f(x)\, dx}$$

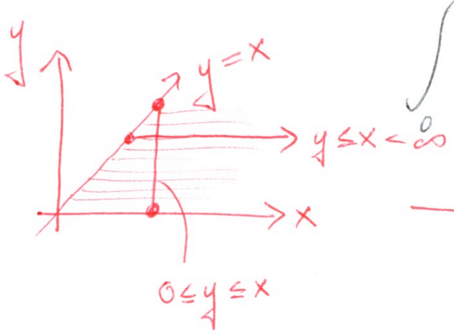
since, in the Riemann sum, $\boxed{f(x)\, \Delta x \approx P(x \leq X \leq x + \Delta x)}$

<u>Prop. 1</u> For a continuous random variable X with pdf $f(x)$, composition with any other (measurable) real-valued function $g : R(X) \subseteq \mathbb{R} \to \mathbb{R}$ yields another random variable $g(X) \equiv g \circ X$, and

$$E[X] = \int_{-\infty}^{\infty} x f(x)\, dx$$

$$\Longrightarrow \boxed{E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\, dx}$$

<u>pf:</u> First, note that for any nonneg. random variable Y we have $E[Y] = \int_0^{\infty} P(Y > y)\, dy$, since



$$\int_0^{\infty} P(Y > y)\, dy = \int_0^{\infty}\int_y^{\infty} f(x)\, dx\, dy$$

$$= \int_0^{\infty}\left(\int_0^{x} 1\, dy\right) f(x)\, dx$$

Fubini's Thm.

$$= \int_0^{\infty} x f(x)\, dx = E[Y]$$

Prop. 3 $\boxed{Var(aX+b) = a^2 Var(X)}$

pf: As in the discrete case: let

$$\nu \overset{def}{=} E[aX+b] = aE[X]+b$$
$$= a\mu + b$$

and observe

$$Var(aX+b) = E[((aX+b)-\nu)^2]$$
$$= E[(aX+\cancel{b}-a\mu-\cancel{b})^2]$$
$$= E[a^2(X-\mu)^2]$$
$$\overset{Cor. 2}{=} a^2 E[(X-\mu)^2]$$
$$= a^2 Var(X)$$

QED

The Cumulative Distribution of a Continuous X

Finally, let us consider the cumulative distribution F of a continuous random variable:

$$\boxed{\begin{aligned} F(x) &= P(X \leq x) \\ &= P(X^{-1}(-\infty, x]) \\ &= \int_{-\infty}^{x} f(t)\, dt \end{aligned}}$$

The relationship between F, the distribution, and f, the density, is (by Fundamental Thm. of Calculus II)

$$\boxed{\begin{aligned} \frac{d}{dx} F(x) &= \frac{d}{dx} \int_{-\infty}^{x} f(t)\, dt \\ &= f(x) \end{aligned}}$$

Thm. 4 Let X be continuous with density f. Then for any $g: \mathbb{R} \to \mathbb{R}$ which is monotonic, the random variable

$$\boxed{Y = g(X)}$$

has probability density

$$\boxed{f_Y(y) = \begin{cases} (f \circ g^{-1})'(g^{-1}(y)) \, |(g^{-1})'(y)|, & y \in R(g) \\ 0, & y \notin R(g) \end{cases}}$$

pf: If g is increasing $(x_1 \le x_2 \Rightarrow g(x_1) \le g(x_2))$ then (assuming, as we do, that g is differentiable, & so continuous) g is invertible, and g^{-1} is also increasing $(y_1 \le y_2 \Rightarrow \underbrace{g^{-1}(y_1)}_{=x_1} \le \underbrace{g^{-1}(y_2)}_{=x_2})$, and similarly g is decreasing $(x_1 \le x_2 \Rightarrow g(x_1) \ge g(x_2))$ implies g^{-1} is decreasing $(y_1 \le y_2 \Rightarrow g^{-1}(y_1) = x_1 \ge x_2 = g^{-1}(y_2))$

If g is increasing, then

$$F_Y(y) = P(Y \le y) = \int_{(-\infty, y]} f_Y(t)\,dt$$

$$= P(g(X) \le y)$$

$$= P(X \le g^{-1}(y)) = \int_{(-\infty, g^{-1}(y)]} f_X(s)\,ds$$

$$= F_X(g^{-1}(y))$$

$$\Longrightarrow \quad \underbrace{\frac{d}{dy} F_Y(y)}_{= f_Y(y)} = \frac{d}{dy}(F_X \circ g^{-1})(y)$$

chain rule

$$= \underbrace{\frac{dF_X}{dx}(g^{-1}(y))}_{= f_X(g^{-1}(y))} \cdot \frac{d}{dy} g^{-1}(y)$$

or

$$f_Y(y) = (f_X \circ g^{-1})(y) \cdot \underbrace{(g^{-1})'(y)}_{\ge 0 \text{ since } g^{-1} \text{ is increasing}}$$

$$= (f_X \circ g^{-1})(y) \left| (g^{-1})'(y) \right|$$

If g is decreasing, then $(g^{-1})'(y) < 0$, &

$$F_Y(y) = P(Y \le y) = P(g(X) \le y)$$

$$= P(X \ge g^{-1}(y))$$

$$= 1 - P(X < g^{-1}(y))$$

$$= 1 + \underbrace{P(X = g^{-1}(y))}_{= \int_{g(y)}^{g(y)} f_X(t)\,dt \;=\; 0} - P(X \leq g^{-1}(y))$$

$$= 1 - F_X(g^{-1}(y))$$

So

$$\frac{d}{dy} F_Y(y) = \frac{d}{dy}\left(1 - (F_X \circ g^{-1})(y)\right)$$

$$= -\frac{d}{dx} F_X(g^{-1}(y)) \cdot \frac{d}{dy} g^{-1}(y)$$

$$= \underbrace{-(f_X \circ g^{-1})(y)}_{\text{take the } -1 \text{ \& put it with } (g^{-1})'} \cdot \underbrace{(g^{-1})'(y)}_{< 0 \text{ since } g \text{ is decreasing}}$$

$$= (f_X \circ g^{-1})(y) \underbrace{|(g^{-1})'(y)|}_{= -(g^{-1})'(y)}$$

QED

ex. 1 Let $\boxed{Y = X^n,}$ then $\boxed{g(x) = x^n.}$

Case 1: n is odd, X arbitrary

If n is odd, then $g'(x) = nx^{n-1} \geq 0$, so g is increasing on $\mathbb{R}$ & the theorem applies.

Case 2: n even, $X \geq 0$

If n is even, then we must instead require X to be nonnegative, so that

~~g(X)~~ $X \geq 0 \implies x \geq 0$ in $g(x) = x^n$;

i.e. $\mathrm{dom}(g) = [0, \infty)$, & on that interval

$g'(x) = nx^{n-1} \geq 0$ bec. $x \geq 0 \implies x^{n-1} \geq 0$.

The theorem therefore applies, g is incr.:

$$P(X \in B) = \int_B f(x)\, dx$$

$$\implies P(g(X) \in B) = \int_B f(g^{-1}(y)) \cdot (g^{-1})'(y)\, dy$$

which in our case, since

$$g^{-1}(y) = y^{1/n} \implies (g^{-1})'(y) = \frac{1}{n} y^{1/n - 1}$$

means

$$\boxed{P(X^n \in B) = \int_{g(B)} f(y^{1/n}) \cdot \frac{1}{n} y^{1/n - 1} \, dy}$$

For example,

$$\boxed{\begin{aligned} P(X^2 \in B) &= \int_{g(B)} \left(f(\sqrt{y}) / 2\sqrt{y} \right) dy \\ P(X^3 \in B) &= \int_{g(B)} \left(f(\sqrt[3]{y}) / 3\sqrt[3]{y^2} \right) dy \end{aligned}}$$

etc.