Properties of cdfs, whether continuous, discrete, etc. r.v.

<u>Prop</u>: A function F is the cdf of some random variable iff:

① F is nondecreasing

② $\lim_{c \to +\infty} F(c) = 1$, $\lim_{c \to -\infty} F(c) = 0$

③ F is right continuous, i.e. $F_X(c) = F_X(c+)$ for all c.

examples: a, c, f valid; others not.

---

for continuous-type random variables, cdf is integral of a function

$$F_X(c) = \int_{-\infty}^{c} f_X(u)\, du$$

where f_X is called the probability density function. (pdf)

The <u>support</u> of pdf f_X is set of u such that $f_X(u) > 0$.

By the fundamental thm of calculus, if f_X is continuous, then

$f_X = F_X'$.

In particular, this implies F_X is continuous.

But if F_X continuous, then no jumps, so for any constant value c,

$Pr[X = c] = 0$.

Instead interpret pdf value as probability that X is near c (in small interval around c), from limit definition of derivative.

Integral determines probability: $Pr[a \leq X < b] = \int_a^b f_X(u)\, du$.

Properties of pdf:

(1) ~~everywhere~~ f_X is nonnegative.

(2) $\int_{-\infty}^{\infty} f_X(u)\,du = 1$, since $= \lim_{a\to-\infty} \lim_{b\to\infty} F_X(b) - F_X(a) \left[= \int_{-\infty}^{\infty} f_X(u)\,du\right] = 1.$

[note a pdf can be arbitrarily large; Ex: $f_X(x) = \begin{cases} 1/2\sqrt{x} & , 0 < x \le 1 \\ 0 & , \text{else} \end{cases}$ is valid since $\int_0^1 \frac{1}{2\sqrt{x}}\,dx = \sqrt{x}\Big|_0^1 = 1$.

Moments of continuous r.v.:

$$\mu_X = E[X] = \int_{-\infty}^{\infty} u f_X(u)\,du$$

LOTUS:

$$E[g(X)] = \int_{-\infty}^{\infty} g(u) f_X(u)\,du.$$

Linearity still holds (same for integrals as sums).

$$E[aX^2 + bX + c] = a\,E[X^2] + b\,E[X] + c.$$

Variance is still $Var[X] = E[X^2] - \mu_X^2$.

Can also consider mixed random variables with discrete and continuous parts.

Represent in pdf using Dirac delta functions:

Exercise: If cdf is $F(x) = \begin{cases} e^{x-2} & , \text{if } x < 1 \\ \frac{1}{2} & , \text{if } 1 \le x < 2 \\ 1 & , \text{if } 2 \le x \end{cases}$

what is pdf?

Maximum of several random variables

You are allowed to do a problem three times and final score is maximum of test scores:

$$X = \max\{X_1, X_2, X_3\}.$$

Assume the score in each test takes value in $\{1, 2, \dots, 10\}$ with equal probability $\frac{1}{10}$, independently of the other tests. What is the pmf of the final score?

We do the computation indirectly. First compute cdf F_X and then obtain the pmf as the first difference:

$$p_X(k) = F_X(k) - F_X(k-1), \quad k = 1, \dots, 10$$

which is the analog of the first derivative.

$$\begin{aligned} F_X(k) &= \Pr[X \le k] \\ &= \Pr[X_1 \le k, X_2 \le k, X_3 \le k] \\ &= \Pr[X_1 \le k]\Pr[X_2 \le k]\Pr[X_3 \le k] \\ &= \left(\frac{k}{10}\right)^3 \end{aligned}$$

Thus

$$p_X(k) = \left(\frac{k}{10}\right)^3 - \left(\frac{k-1}{10}\right)^3, \quad k = 1, 2, \dots, 10.$$