Class Notes 2-20-2019

- **Properties of CDFs:**

  (i) **Monotonicity:** $F(s) \leq F(t)$ for all $s < t$.

  (ii) **Right continuity:** $F(t) = \lim_{s \searrow t} F(s) = \lim_{s \to t^+} F(s) = F(t^+)$ for all $t \in \mathbb{R}$

  (iii) **Limits:** $\lim_{t \to -\infty} F(t) = 0$ and $\lim_{t \to \infty} F(t) = 1$

- $P(X \leq x) = F(x)$ **and** $P(X < x) = F(x-)$**:**

$$P(X < x) = \lim_{s \nearrow x} F(s) = \lim_{s \to x^-} P(X \leq s) = F(x-)$$

The proof of this and the previous two limits follows from (Finer Points Fact 1.39)

$$A_1 \subset A_2 \subset ... \subset A_n \subset ... \quad \text{with} \;\; A = \cup_{i=1}^{\infty} A_i \Rightarrow P(A_n) \to P(A) \;\; \text{as } n \to \infty$$

or its equivalent (by complement) form

$$B_1 \supset B_2 \supset ... \supset B_n \supset ... \quad \text{with} \;\; B = \cap_{i=1}^{\infty} B_1 \Rightarrow P(B_n) \to P(B) \;\; \text{as } n \to \infty$$

This limit behavior is a direct consequence of axiom iii) of probability measures.

$$\cup_{n=1}^{\infty} \left(-\infty, x - \frac{1}{n}\right] = (-\infty, x) \quad \Rightarrow \quad P(X \leq x - 1/n) = F(x - 1/n) \to P(X < x) = F(x-)$$

In olden times it was custom to define the cdf via $P(X < x)$ east of the iron curtain and as $P(X \leq x)$ to the west. That may not have changed. Thus you need to watch out what definition is used. $PX < x)$ is left continuous in x. They used to joke about this. Note that

$$F(x) - F(x-) = P(X \leq x) - P(X < x) = P(X = x) \quad = \quad \text{size of the jump in } F \text{ at } x$$

- **Example: CDF of** $f(x) = x^{-2} I_{[1,\infty)}(x)$

$$F(s) = \int_{-\infty}^{s} f(x)dx = \int_{1}^{s} \frac{1}{x^2} dx = \left(-\frac{1}{x}\right)\Bigg|_1^s = 1 - \frac{1}{s} \quad \text{for } s \geq 1 \text{ and } \; F(s) = 0 \quad \text{for } s < 1$$

- **Dartboard Example Revisited:** Consider a disk with radius r_0. A dart is thrown randomly at it with an assumed uniform distribution for its point of impact. Define as random variable its distance R from the center of the disk. Find its cdf and density.

$$F_R(t) = P(R \leq t) = \frac{\pi t^2}{\pi r_0^2} = \frac{t^2}{r_0^2} \;\; \text{for} \;\; 0 \leq t \leq r_0, \;\; F_R(t) = 0 \;\; \text{for} \;\; t < 0 \;\; \text{and} \;\; F(t) = 1 \;\; \text{for} \;\; T > r_0$$

$$f_R(t) = \frac{2t}{r_0^2} I_{[0,r_0]}(t)$$

- **Examples where** X **is neither Discrete nor Continuous:**
  Peter has an insurance policy with a \$ 500 deductible, i.e., he pays any claims up to \$ 500, any amount above is covered by the insurance. Suppose that the damage from the next accident is a random variable Y uniformly distributed over $[100, 2000]$. Let X be the random variable

which is the amount that Peter pays in the case of such an accident. What is the distribution function of X. Clearly $X = Y$ if $Y < 500$ and $X = 500$ otherwise. We have

$$F_X(x) = P(X \leq x) = P(Y \leq x) = \frac{x - 100}{1900} \quad \text{for } 100 \leq x < 500 \quad \text{the continuous part}$$

$$P(X = 500) = P(Y \geq 500) = \frac{2000 - 500}{1900} = \frac{15}{19} \quad \text{the discrete part}$$

- **Expectation (Discrete Case):** The expectation or mean of a discrete random variable X is defined as

$$E(X) = \sum_k k p_X(k) \quad \text{with summation over all possible values } k \text{ of } X$$

It is also called the first moment of X and is commonly denoted by $\mu = \mu(X) = \mu_X = E(X)$. It can be viewed as the center of gravity of the possible values of X when these are loaded with weights equal to the corresponding probabilities. If a denoted the location of a fulcrum among the values of X endowed with such probability weights then we have balance when

$$\sum_{k \leq a} (a - k) p_X(x) = \sum_{k > a} (k - a) p_X(x) \quad \left(\Longleftrightarrow \sum_{k < a} (a - k) p_X(x) = \sum_{k > a} (k - a) p_X(x) \right)$$

the moments on the left balance out the moments on the right. This happens when $a = E(X)$

$$0 = \sum_{k \leq a} (a - k) p_X(k) - \sum_{k > a} (k - a) p_X(k) = a \sum_k p_X(k) - \sum_k k p_X(k) = a - \sum_k k p_X(k)$$

- **Gambling:**
When placing bets in games of chance the payouts or losses are discrete random variables and the expected value represents the expected payout or loss (if negative) in the long run. Casinos make sure that it is always slightly negative, i.e., in their favor. The phrase in the long run will become more meaningful later. On how to beat the house in roulette see `https://en.wikipedia.org/wiki/Eudaemons` or read the entertaining book The Eudaemonic Pie by Thomas A Bass.

- **Mean of a Binomial Random Variable:**

$$k \binom{n}{k} = n \binom{n-1}{k-1} \qquad E(X) = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k}$$

$$\text{using } j = k - 1 \qquad = np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{n-1-j} = np$$

The mean is linearly increasing in n and in p, intuitively appealing.

- **Mean of a Bernoulli Random Variable:**
This is the simplest it can get (except for constant RVs).
X has values 0 and 1 with $P(X = 1) = 1$ and $P(X = 0) = 1 - p$.

$$E(X) = 0 \cdot (1 - p) + 1 \cdot p = p$$

A particular case of a Bernoulli RV is the indicator random variable of an event $B \subset \Omega$, denoted by

$$I_B(\omega) = 1 \quad \text{if} \quad \omega \in B \quad \text{and} \quad I_B(\omega) = 0 \quad \text{if} \quad \omega \in B^c \ \text{or} \ \omega \notin B \quad \Rightarrow \quad E(I_B) = P(B)$$

- **Expectation of a Geometric Random Variable:**
  $P(X = k) = p(1-p)^{k-1} = pq^{k-1}$ for $k = 1, 2, \ldots$, where $q = 1 - p$. Then

$$E(X) = \sum_{k=1}^{\infty} kpq^{k-1} = p \sum_{k=0}^{\infty} \frac{d}{dq} q^k = p \frac{d}{dq} \sum_{k=0}^{\infty} q^k = p \frac{d}{dq} \frac{1}{1-q} = p \frac{1}{(1-q)^2} = \frac{1}{p}$$

  which is inversely linear in p, also intuitively appealing.

- **Expectation of Continuous Random Variables:**
  If a continuous RV X has density $f(x)$ we define its expectation as

$$E(X) = \mu = \int_{-\infty}^{\infty} x f(x) dx = \int_{-\infty}^{\infty} y f(y) dy$$

  The x and y in the integration are just dummy variables and can be any (non-confusing) letter symbol. The X in $E(X)$ is the name of the random variable used in this instance. We have been using that symbol X a lot in naming all kinds of RVs.
  The above definition is very analogous to the summation in the discrete case if we view the $\int$ as a summation of infinitesimals over the continuum of $\mathbb{R}$.

- **Mean of a Uniform Random Variable** $X \sim \text{Unif}[a, b]$**:**

$$E(X) = \int_{-\infty}^{\infty} x \frac{1}{b-a} I_{[a,b]}(x) dx = \int_a^b x \frac{1}{b-a} dx = \left(\frac{1}{b-a} \frac{x^2}{2} \right) \Bigg|_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}$$

  again an intuitive center of gravity or balance.