Class Notes 2-8-2019

- **Infinitesimals or Probability Interpretation of Densities:**
  $f(x)$ is not the probability $P(X = x) = 0$. However, when f is continuous at a point a then

  $$P(a - \epsilon/2 < X < a + \epsilon/2) \approx \epsilon f(a) \quad \text{for small } \epsilon > 0$$

  because

  $$P(a - \epsilon/2 < X < a + \epsilon/2) = \int_{a-\epsilon/2}^{a-\epsilon/2} f(x)\,dx \approx \epsilon f(a)$$

  If f is right continuous at a we have

  $$P(a < X < a + \epsilon) \approx \epsilon f(a)$$

  with a corresponding statement when left continuity holds at a.

- **Example of Infinitesimal Approximation:**
  Consider the density $f(x) = 3x^2 I_{[0,1]}(x)$ and compare the approximation with the exact

  $$P(0.5 < X < .51) = \int_{.5}^{.51} 3x^2\,dx = x^3\Big|_{.5}^{.51} = .51^3 - .5^3 = .007651$$

  while the approximation gives $3(.5)^2 \times .01 = .0075$, not bad.

- **Inverse Use of Infinitesimal Method for Finding $f(x)$, Dartboard Example:**
  Consider throwing a dart at a dartboard with radius 9 inches. Assume that the dart land at point that is chosen uniformly at random on this board. Let R be the distance of the dart from the center of the board. We will use the approximation $P(t < R < t + \epsilon)/\epsilon \approx f_R(t)$ to find $f_R(t)$. The event $t < R < t + \epsilon$ states that the chosen dart landing point is with in the annulus with radii t and $t + \epsilon$ which has area

  $$\pi(t + \epsilon)^2 - \pi t^2 = 2\pi t\epsilon + \pi\epsilon^2$$

  Since the dart location is chosen uniformly at random we get as the probability of landing in the annulus its area over the area of the full circle, i.e.,

  $$P(t < R < t + \epsilon) = \frac{2\pi t\epsilon + \pi\epsilon^2}{\pi 9^2} = \frac{2t\epsilon}{9^2} + \frac{\epsilon^2}{9^2}$$

  thus for $0 < t < 9$

  $$P(t < R < t + \epsilon)/\epsilon = \frac{2t}{9^2} + \frac{\epsilon}{9^2} \longrightarrow \frac{2t}{9^2} = f_R(t) \quad \text{as } \epsilon \to 0$$

  For $t = 0, 9$ define $f_R(t)$ arbitrarily, via extension by continuity, e.g., $f_R(0) = 0$ and $f_R(9) = 2/9$. $f_R(t) = 0$ for $t > 9$ or $t < 0$. Check that $f_R(t)$ is a density.

- **Where is the Sample Space Ω?**
  So far when dealing with p.d.f.'s and random variables we have started with Ω twice, a uniformly random choice over an interval or over a circle. Most of the time we will start out with specifying a density of a random variable X and neglect Ω. But it can always be constructed from $f(x)$. Namely take $\Omega = \mathbb{R}$ and assign probabilities to $B \subset \Omega = \mathbb{R}$ via

$P(B) = \int_B f(x)dx$ and define the random variable as a function $X : \Omega \to \mathbb{R}$ via $X(\omega) = \omega$ for any $\omega \in \Omega$. This would then agree with

$$P(X \in B) = P(\{\omega : X(\omega) = \omega \in B\}) = P(B) = \int_B f(x)dx$$

Here B can be any interval or countable collection of intervals. Thus we will most often ignore the sample space Ω that (trivially) exists in the background.

- **Cumulative Distribution Function (C.D.F.)**
  This is another way of describing the distribution of a random variable. It is a unifying way of doing so, whether the random variable is discrete or continuous (or even neither).
  The cumulative distribution function (c.d.f.) of a random variable X is defined as

$$F(s) = P(X \leq s) \quad \text{for all } s \in \mathbb{R}$$

  From this we get

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$$

  Be mindful of the use of $<$ and $\leq$.
  Knowing these probabilities determines the distribution of X.

- **The C.D.F. of a Discrete Random Variable:**

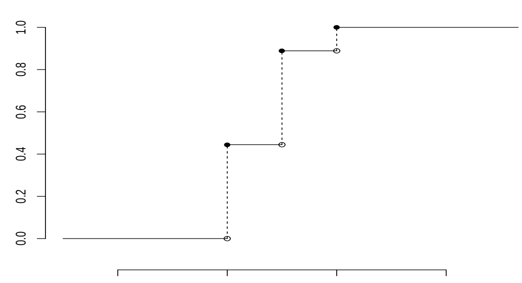
$$F(s) = P(X \leq s) = \sum_{k:k \leq s} P(X = k) = \sum_{k \leq s} P(X = k) = \sum_{k \in B} P(X = k) = P(X \in B)$$

  with $B = (-\infty, s]$.

- **Binomial Example:**
  Let $X \sim \text{Bin}(2, 1/3)$ then

$$P(X = 0) = \binom{2}{0}\left(\frac{1}{3}\right)^0\left(\frac{2}{3}\right)^2 = \frac{4}{9}, \;\; P(X = 1) = \frac{4}{9}, \;\; P(X = 2) = \frac{1}{9}$$

$$F(s) = \begin{cases} 0, & s < 0 \\ \frac{4}{9}, & 0 \leq s < 1 \\ \frac{8}{9}, & 1 \leq s < 2 \\ 1, & s \geq 2 \end{cases}$$



The graph of the CDF jumps at every possible value k of X with positive probability $P(X = k) > 0$ and it is flat in between and beyond. The size of the jump at k is $P(X = k)$.

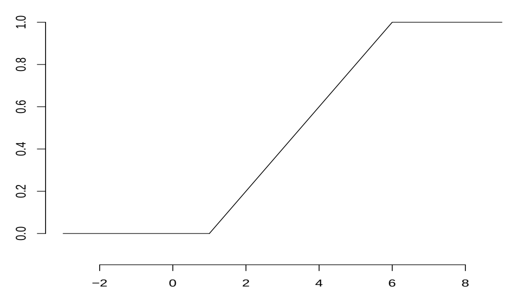
- **C.D.F. of a Continuous Random Variable:**
For a continuous random variable with density f we have

$$F(s) = P(X \leq s) = P(X \in B) = \int_B f(x)dx = \int_{-\infty}^{s} f(x)dx \quad \text{with } B = (-\infty, s]$$

- **Uniform Example:**
Let $X \sim U(1, 6)$ then $f(x) = 1/5$ for $x \in [1, 6]$ and $f(x) = 0$ else. For $s \in [1, 6]$ we have

$$F(s) = \int_{-\infty}^{s} f(x)dx = \int_{1}^{s} \frac{1}{5}dx = \frac{s-1}{5}$$

$F(s) = 0$ for $s < 1$ and $F(s) = 1$ for $s > 6$.



The graph of the CDF has no jumps. It is continuous, which is the reason for calling random variables with densities continuous.

- **Finding the PMF or PDF from the CDF:**
  - If F is piecewise constant, then X is a discrete random variable with the jump sizes at $k_1, k_2, \ldots$ representing the $P(X = k_i)$.
  - If F is continuous and has a derivative on the real line except at finitely many points, then X is a continuous random variable with pdf $f(x) = F'(x)$. At points where the derivative does not exist set $f(x)$ to arbitrary values (≥ 0).

- **Examples:**
  - If $F(x) = 0, 1/3, 1/2, 3/4, 1$ as $x < 1$, $1 \leq x < 2$, $2 \leq x < 4$, $4 \leq x < 7$, $7 \leq x$, then $p(1) = 1/3$, $p(2) = 1/2 - 1/3 = 1/6$, $p(4) = 3/4 - 1/2 = 1/4$ and $p(7) = 1 - 3/4 = 1/4$. $p(x) = 0$ for all other x. As cross-check note $1/3 + 1/6 + 1/4 + 1/4 = 1$.
  - If $F(x) = 0, x/3, (2x - 1)/3, 1$ as $x < 0, 0 \leq x < 1, 1 \leq x < 2, 2 \leq x$ then $f(x) = \frac{1}{3}I_{(0,1)}(x) + \frac{2}{3}I_{(1,2)}(x)$ where we set $f(x) = 0$ at the points $x = 0, 1, 2$ where the derivative of F does not exist. Note that the total area under $f(x)$ is 1.

- **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 $\lim_{t \to \infty} F(t) = 1$