MA 6.101
Probability and Statistics

Tejas Bodas

Assistant Professor, IIIT Hyderabad

Geometric random variable

- Consider a biased coin (head with probability p) and suppose you keep tossing it till head appears the first time.
- Let random variable N denote the number of tosses needed for head to appear first time.
- What is the PMF of N? $p_N(k) = (1-p)^{k-1}p$.
- HW: What is $E[N], E[N^2], Var(N)$?
- What is $\bar{F}_N(k) := 1 - F_N(k) = P(N > k)$?
- What is $P(N > k + m \mid N > k)$? $(= P(N > m))$ (memoryless property)

Poisson random variable

- A Poisson random variable X comes with a parameter λ and has $\Omega' = \mathbb{Z}_{\geq 0}$
- PMF: $p_X(k) = e^{-\lambda}\frac{\lambda^k}{k!}$
- Intuitively its a limiting case of the Binomial distribution with n increasing and p decreasing such that np converges to λ.
- Mean of binomial is np so p should decrease while n increases.
- Read the Wiki page on Poisson limit theorem.
- We will see more of this when we see Poisson Prcoesses.

Induced measure P_X and CDF

- The cumulative distribution function (CDF) $F_X(x)$ can be expressed using induced measure P_X.
- Suppose the domain of P_X is $\mathcal{B}(\mathbb{R})$. $\mathcal{B}(\mathbb{R})$ is made up of sets of the form $(-\infty, a]$ for $a \in \mathbb{R}$.
- $P_X((-\infty, x]) = \mathbb{P}\{w \in \Omega : X(w) \leq x\} := F_X(x)$.
- This is a general definition of CDF (applicable for both continuous or discrete).
- If $F_X(\cdot)$ is continuous (resp. piecewise continuous), then X is continuous (resp. discrete) random variable.

For a r.v. X, its CDF satisfies the following

- $F_X(\infty) = 1$ and $F_X(-\infty) = 0$ when $P(-\infty < X < \infty) = 1$.

- $F_X : \mathbb{R} \to [0, 1]$ is non-decreasing and right continuous.

- At point of discontinuity x we have
  1. right hand limit $F_X(x+) := \lim_{\epsilon \downarrow 0} F_X(x + \epsilon)$
  2. left hand limit $F_X(x-) := \lim_{\epsilon \uparrow 0} F_X(x - \epsilon)$
  3. $F_X(x-) \neq F_X(x+)$.
  4. $F_X(x)$ could be set to either of the two. Which one?

- Right continuity mandates that at point of discontinuity, we have $F_X(x) = F_X(x+)$.

- By default, $F_X(x) = F_X(x+) = F_X(x-)$ if $F_X(x)$ is continuous at x.

Right-continuity

$F_X : \mathbb{R} \to [0, 1]$ is non-decreasing and right continuous.

Proof

- Consider $a < b$ where a and b are arbitrary. We want to show that $F_X(a) \le F_X(b)$.
- Define $A := \{\omega \in \Omega : X(\omega) \le a\}$, $B := \{\omega \in \Omega : X(\omega) \le b\}$.
- Easy to see that $A \subseteq B$ and hence $\mathbb{P}(A) \le \mathbb{P}(B)$.
- $F_X(a) = P_X((-\infty, a]) = \mathbb{P}(A) \le \mathbb{P}(B) = F_X(b)$.
- This proves the non-decreasing part.

Right-continuity

$F_X : \mathbb{R} \to [0,1]$ is non-decreasing and right continuous.

Proof for right-continuity

- We want to prove that $F_X(x) = F_X(x+)$.
- Consider a sequence of numbers $\{x_n\}$ decreasing to x. In this case, we have $F_X(x+) = \lim_{x_n \downarrow x} F_X(x_n)$.
- Define $A_n := \{\omega : X(\omega) \leq x_n\}$ and $A := \{\omega : X(\omega) \leq x\}$.
- Is $A_n \uparrow A$ or $A_n \downarrow A$? Clearly, $A_n \downarrow A$.
- From continuity of probability, $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A)$.
- This implies $\lim_{x_n \downarrow x} F_X(x_n) = F_X(x)$. $\square$
- You cannot prove the other way by considering $x_n \uparrow x$ because $\cup_n(-\infty, x_n] = (-\infty, x)$ and $P_X(-\infty, x) \neq F_X(x)$.

Continuous random variables

Continuous random variables

- A random variable defined on $\mathbb{R}$ is discrete, if $F_X(\cdot)$ is piecewise constant.
- A random variable defined on $\mathbb{R}$ is continuous, if $F_X(\cdot)$ is a continuous function.
- Examples of Continuous random variables
  1. Pick a number uniformly from $[a, b]$.
  2. Time interval between successive customers entering DMart.
  3. Travel time from office to home.
  4. Level of water in a dam or pending workload on a server.

Continuous random variables

- Associated with a continuous random variable is a probability density function (pdf) $f_X(x)$ for all $x \in \mathbb{R}$. Its unit is probability per unit length and is defined as

$$
\begin{aligned}
f_X(x) \quad &:= \quad \lim_{\Delta \to 0^+} \frac{P(x < X \leq x + \Delta)}{\Delta} \\
&= \quad \lim_{\Delta \to 0^+} \frac{F_X(x + \Delta) - F_X(x)}{\Delta} \\
&= \quad \frac{dF_X(x)}{dx}.
\end{aligned}
$$

- Alternatively, a random variable X is continuous if there exists a non-negative real valued probability density function (PDF) $f_X(\cdot)$ such that $F_X(x) = \int_{u=-\infty}^{x} f_X(u)du$.

$$\boxed{\tfrac{dF_X(x)}{dx} = f_X(x) \text{ or } P_X(x < X \leq x + h) \simeq f_X(x)h.}$$

Properties of pdf

- $P_X(\mathbb{R}) = \int_{u=-\infty}^{\infty} f_X(u)du = 1.$
- $P_X(a \leq X \leq b) = \int_a^b f_X(u)du.$ (Area under the curve)
- In general, $P_X(B) = \int_{u \in B} f_X(u)du.$
- $P_X(a \leq X \leq b) = P_X(a < X < b) = P_X(a \leq X < b) = P_X(a < X \leq b)$
- $P_X(X = a) = 0.$ (no mass at any point)

Mean, Variance, Moments

- $E[X] = \int_{-\infty}^{\infty} u f_X(u) du$
- $E[X^n] = \int_{-\infty}^{\infty} u^n f_X(u) du$
- $E[g(X)] = \int_{-\infty}^{\infty} g(u) f_X(u) du$
- $Var[X] = E[g(X)]$ where $g(x) = (x - E[X])^2$.
- For $Y = aX + b$, $E[Y] = aE[X] + b$.
- For $Y = aX + b$, $F_Y(y) = F_X(\frac{y-b}{a})$ when $a \geq 0$.
- For $Y = aX + b$ and $a < 0$, $F_Y(y) = 1 - F_X(\frac{y-b}{a})$.