MA 6.101
Probability and Statistics

Tejas Bodas

Assistant Professor, IIIT Hyderabad

Logistics

- No tutorial tomorrow!
- This means, you are on your own for the Quiz next week!!
- We can have a TA hours (90 min) on 23rd.
- How about 3:30pm to 5pm on 23th (Tuesday)?

Motivation to random variables

Random variable

- Given a random experiment with associated $(\Omega, \mathcal{F}, \mathbb{P})$, it is sometimes difficult to deal directly with $\omega \in \Omega$. eg. rolling a dice ten times.
- Notice that each sample point $\omega \in \Omega$ is not a number but a sequence of numbers.
- Also, we may be interested in functions of these sample points rather than samples themselves. eg: Number of times 6 appears in the 10 rolls.
- In either case, it is often convenient to work in a new *simpler* probability space rather than the original space.
- Random variable is a device which precisely helps us make this mapping from $(\Omega, \mathcal{F}, \mathbb{P})$ to a 'simpler' $(\Omega', \mathcal{F}', P_X)$.
- P_X is called as an induced probability measure on Ω'.

Random variable as a measurable function

A random variable X is a function $X : \Omega \to \Omega'$ that transforms the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to $(\Omega', \mathcal{F}', P_X)$ and is '$(\mathcal{F}, \mathcal{F}')$-measurable'.

- The map $X : \Omega \to \Omega'$ implies $X(\omega) \in \Omega'$ for all $\omega \in \Omega$.
- A random variable could be non-injective and non-surjective.
- For event $B \in \mathcal{F}'$, the pre-image $X^{-1}(B)$ is defined as $X^{-1}(B) := \{\omega \in \Omega : X(\omega) \in B\}$

The '$(\mathcal{F}, \mathcal{F}')$-measurability' implies that for every $B \in \mathcal{F}'$, we have $X^{-1}(B) \in \mathcal{F}$.

Random variable as a measurable function

The '$(\mathcal{F}, \mathcal{F}')$-measurability' implies that for every $B \in \mathcal{F}'$, we have $X^{-1}(B) \in \mathcal{F}$.

- Since $X^{-1}(B) \in \mathcal{F}$, it can be measured using $\mathbb{P}$.
- What is $P_X(B)$?
- $P_X(B) := \mathbb{P}(X^{-1}(B))$ for all $B \in \mathcal{F}'$.
- $P_X(B)$ is therefore called as the induced probability measure.
- What if there is no $\omega \in \Omega$ such that $X(\omega) \in B$?

Random variables

- In general, the following convention is followed in most books:
  - Ω' will be the set of real numbers, denoted by $\mathbb{R}$.
  - $\mathcal{F}'$ as a result will be Borel σ-algebra, denoted by $\mathcal{B}(\mathbb{R})$.
  - Remember $\mathcal{B}(\mathbb{R})$?

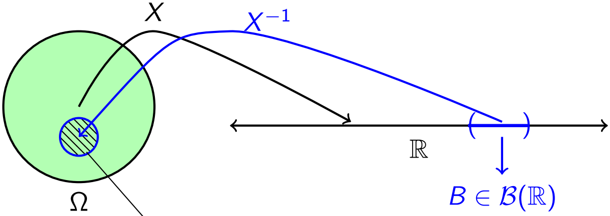
Borel σ-algebra

- Borel σ-algebra $\mathcal{B}(\mathbb{R})$:
  If $\Omega = \mathbb{R}$, then $\mathcal{B}(\mathbb{R})$ is the event set generated by open sets of the form (a, b) where $a \leq b$ and $a, b \in \mathbb{R}$.

- $\mathcal{B}(\mathbb{R})$ contains intervals of the form

$$
\begin{gathered}
[a, b] \\
[a, b) \\
(a, \infty) \\
[a, \infty) \\
(-\infty, b] \\
(-\infty, b) \\
\{a\}
\end{gathered}
$$

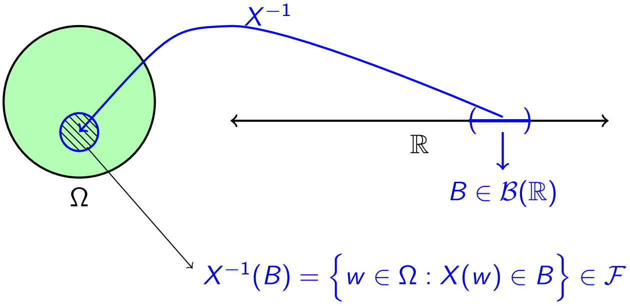
Random variables ($\Omega' = \mathbb{R}$)



$$X^{-1}(B) := \Big\{ w \in \Omega : X(w) \in B \Big\} \in \mathcal{F}$$

- $\Omega \xrightarrow{X} \mathbb{R}$, $\mathcal{F} \xrightarrow{X} \mathcal{B}(\mathbb{R})$, and $P(.) \xrightarrow{X} P_X(.)$
- Care must be taken such that the events you consider in the new event space $\mathcal{B}(\mathbb{R})$ are also valid events included in $\mathcal{F}$.
- $X^{-1}(B)$ is called as the preimage or the inverse image of B.

Definition of a random variables



A random variable X is a map $X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$ such that for each $B \in \mathcal{B}(\mathbb{R})$, the inverse image $X^{-1}(B) := \{w \in \Omega : X(w) \in B\}$ satisfies

$$X^{-1}(B) \in \mathcal{F} \text{ and}$$
$$P_X(B) = \Pr(w \in \Omega : X(w) \in B)$$

Random variable

- If Ω' is countable, then the random variable is called a discrete random variable.
- In this case it is convenient to use $\mathcal{F}'$ as power-set.
- If $\Omega' \subseteq \mathbb{R}$ or uncountable, then the random variable is a continuous random variable.
- In this case, $\mathcal{F}' = \mathcal{B}(\mathbb{R})$ and the definition is a bit tricky. We will deal with it later.
- You can also use $\Omega' = \mathbb{R}$ for a discrete random variable and survive! Lets not get into that.
- Notation: Random variables denoted by capital letters like X, Y, Z etc. anf their realizations by small letters x, y, z..

Discrete random variables

Example of rolling two dice

- Example of rolling two dice where we are interested in the sum of two dice.
- Suppose $X =$ sum of two dice. Then we have

$$\begin{aligned}\Omega = \Big\{&(1,1),(1,2),\dots,(1,6)\\ &(2,1),(2,2),\dots,(2,6)\\ &\qquad\quad\vdots \qquad\qquad\qquad\qquad \xrightarrow{X} \qquad \Omega' = \big\{2,3,\dots,12\big\}\\ &(6,1),(6,2),\dots,(6,6)\Big\}\end{aligned}$$

- $\mathcal{F}$ and $\mathcal{F}'$ are power sets of Ω and Ω' respectively.
- Is X $(\mathcal{F},\mathcal{F}')$-measurable?

Example of rolling two dice

- Example of rolling two dice where we are interested in the sum of two dice.
- Suppose X = sum of two dice. Then we have

$$\Omega = \Big\{(1,1),(1,2),\dots,(1,6) \quad (2,1),(2,2),\dots,(2,6) \quad \vdots \quad (6,1),(6,2),\dots,(6,6)\Big\} \xrightarrow{X} \Omega' = \big\{2,3,\dots,12\big\}$$

- $X = 3$ is an event in $\mathcal{F}'$. What is its probability $P_X(3)$?
- $P_X(3) = \mathbb{P}(\{\omega \in \Omega : X(\omega) = 3\}) = \mathbb{P}(\{(1,2),(2,1)\})$.

In general for $x \in \Omega'$, we have $P_X(x) := \mathbb{P}(\{\omega \in \Omega : X(\omega) = x\})$.
Find $P_X(x)$ for all $x \in \Omega'$?

Sum of two dice

- $\Omega' = \{2, 3, \ldots, 12\}$
- $\mathcal{F}' = \mathcal{P}(\Omega)$
- $P_X(x) = \begin{cases} \frac{x-1}{36} & \text{for } x \in \{2, 3, \ldots, 7\} \\ \frac{13-x}{36} & \text{for } x \in \{8, 9, \ldots, 12\}. \end{cases}$
- $Z =$Sum of 4 rolls ? Ω for 4 rolls is even complicated.
- This is where X is useful. $P(Z = 4) = P(X_1 = 2, X_2 = 2)$
- Here X_1 and X_2 are independent copies of random variable X.

> The function $P_X(x)$ for $x \in \Omega'$ is called as a probability mass function of random variable X.

PMF and CDF

- The function $P_X(x)$ for $x \in \Omega$ is called as a probability mass function.
- The cumulative distribution function (CDF) $F_X(\cdot)$ is defined as $F_X(x_1) := \sum_{x \leq x_1} P_X(x) = \mathbb{P}\omega \in \Omega : X(\omega) \leq x_1$.
- What is the CDF for the random variable corresponding to the coin toss or dice experiment?

Expectation and Moments

- How do you define the arithmatic mean of a collection of numbers?
- The mean or expectation of a random variable X is denoted by $E[X]$ and is given by $E[X] = \sum_{x \in \Omega'} x P_X(x)$.
- What is $E[X]$ for the random variable X that corresponds to the outcome of coin toss or dice experiment?
- The n^{th} moment of a random variable X is denoted by $E[X^n]$ and is given by $E[X^n] = \sum_{x \in \Omega'} x^n P_X(x)$.
- For a function $g(\cdot)$ of a random variable X, its expectation is given by $E[g(X)] := \sum_{x \in \Omega'} g(x) P_X(x)$

Recap

- Random variable as a map from $(\Omega, \mathcal{F}, \mathbb{P})$ to $(\Omega', \mathcal{F}', P_X)$
- $(\mathcal{F}\mathcal{F}')-$measurability and induced measure P_X
- Discrete random variables
- PMF and CDF (non-decreasing)
- Expectation and Moments

What's in the name?

- Why call it random variable when it is a deterministic function of $\omega \in \Omega$?
- You are interested in outcomes of experiments which are random.
- You cannot say which $\omega \in \Omega$ is realized and hence cannot say apriori what value X will take.
- X is a variable because each time the experiment is performed, it can take different values $x' \in \Omega'$.
- There is no pattern in the values it can take, hence random.
- PMF goes one step ahead in capturing this randomness in X and assigns a probability to every value $x \in \Omega'$.

Clarification on Notation

- For a discrete random variable, we abused notation and used P_X for both the probability measure as well as the PMF.
- To avoid confusion we will use the notation $p_X(\cdot)$ to denote the PMF. We restate the definition here.
- The probability mass function of a discrete random variable X is defined as $p_X(x) := \mathbb{P}(\{\omega \in \Omega : X(\omega) = x\})$ for $x \in \Omega'$.

Consistency of the PMF

- PMF: $p_X(x) = \mathbb{P}(\{\omega \in \Omega : X(\omega) = x\})$ for $x \in \Omega'$.
- How do you check if p_X is legitimate PMF?
- $\sum_{x \in \Omega'} p_X(x) = 1$. Can you prove this?

$$
\begin{aligned}
\sum_{x \in \Omega'} p_X(x) &= \sum_{x \in \Omega'} \mathbb{P}(\{\omega \in \Omega : X(\omega) = x\}) \\
&= \mathbb{P}(\cup_{x \in \Omega'} \{\omega \in \Omega : X(\omega) = x\}) \\
&= \mathbb{P}(\Omega) \quad \square
\end{aligned}
$$

Linearity of Expectation

- Recall that $E[X] = \sum_{x \in \Omega'} x p_X(x)$.
- Functions of random variables are random variables.
- Furtermore, $E[g(X)] := \sum_{x \in \Omega'} g(x) P_X(x)$
- For $Y = aX + b$, what is $E[Y]$?

$$\begin{aligned} E[Y] &= \sum_{x \in \Omega'} (ax + b) p_X(x) \\ &= a \sum_{x \in \Omega'} x p_X(x) + b \\ &= aE[X] + b. \end{aligned}$$

- What is the PMF of Y?

PMF of Y where $Y = aX + b$.

- Suppose the range of X is $\Omega' = \{x_1, x_2, \ldots, x_n\}$. Then what is the range Ω'' of Y ?

- $\Omega''\{y_1, \ldots, y_n\}$ where $y_i = ax_i + b$ for $i \in \{1, 2, \ldots, n\}$.

- It is easy to see that, $p_Y(y_i) = p_X(x_i)$ for $i \in \{1, 2, \ldots, n\}$.

$$\begin{aligned} E[Y] &= \sum_{y \in \Omega''} y p_Y(y) \\ &= \sum_{x \in \Omega'} (ax + b) p_y(ax + b) \\ &= \sum_{x \in \Omega'} (ax + b) P_x(x) \\ &= aE[X] + b. \end{aligned}$$

- What if $Y = g(X)$ where the function $g(.)$ is many to one? What is the PMF of Y then ?

Towards Variance ..

- Recall $E[X] = \sum_{x \in \Omega'} x p_X(x)$.
- Furthermore, $E[X^n] = \sum_{x \in \Omega'} x^n p_X(x)$.
- In general, $E[g(X)] := \sum_{x \in \Omega'} g(x) p_X(x)$
- Now consider $g(X) = (X - E[X])^2$. $g(X)$ quantifies the square of the deviation of X from the mean.
- Note $g(X)$ cannot track if the deviation is positive or negative!
- $E[g(X)]$ would then tell us the mean of the sqaure of the deviation.
- In fact, $\sqrt{E(g(X))}$ quantifies the deviation.

Variance

- $E[g(X)] = E[(X - E[X])^2]$ is called as the variance of random variable X.
- $Var(X) := E[(X - E[X])^2]$
- HW: Prove that $E[(X - E[X])^2] = E[X^2] - E[X]^2$
- $\sigma_X = \sqrt{Var(X)}$ is called as the standard deviation of X.
- For a fiar coin toss, instead of $\Omega' = \{1, -1\}$, what if we use $\{+100, -100\}$? The latter has more variance!

Examples of discrete random variables

Indicator random variable

- Indicator random variable $1_A(\omega) = \begin{cases} 1, & \text{If } \omega \in A \subseteq \Omega \\ 0, & \text{otherwise.} \end{cases}$
- Its PMF is $p_{1_A}(x) = \begin{cases} \mathbb{P}(A), & \text{when } x = 1 \\ 1 - \mathbb{P}(A), & \text{when } x = 0. \end{cases}$
- This is a discrete random variable even though Ω could be continuous.
- For example, Event A could be that the number picked uniformly on the real line is positive.
- What is its CDF and mean denoted by $E[1_A]$?
- What about its mean variance and moments?

Bernoulli random variable

- Bernoulli random variable $X = \begin{cases} 1, & \text{with probability } p \\ 0, & \text{otherwise.} \end{cases}$
- This is same as an indicator variable but here we do not specify A.
- As a matter of convenience, we will start ignoring Ω from now on.
- These random variables are used in Binary classification in ML. $X = 1$ if image has a cat.
- Basic models of Multi-arm bandit problem assume Bernoulli Bandits.
- $E[X] = p, E[X^n] = p$.

Binomial $B(n, p)$ random variable.

- Consider a biased coin (head with probability p) and toss it n times.
- Denote head by 1 and tail by 0.
- Let random variable N denote the number of heads in n tosses.
- PMF of N?. $P_N(k) = \binom{n}{k} p^k (1-p)^{n-k}$.
- HW: What is $E[N], E[N^2], Var(X)$?

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 $E[N], E[N^2], Var(X)$?

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.