Recap

Recap

- Random variable as a map from $(\Omega, \mathcal{F}, \mathbb{P})$ to $(\Omega', \mathcal{F}', P_X)$

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

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

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)

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?

What's in the name?

- Why call it random variable when it is a deterministic function of $\omega \in \Omega$?

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.

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.

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 $x' \in \Omega'$.

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.

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

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.

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.

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 $x \in \Omega'$.

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?

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?
- $\sum_{x \in \Omega'} p_X(x) = 1$.

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?

$$\sum_{x \in \Omega'} p_X(x) \quad = \quad \sum_{x \in \Omega'} \mathbb{P}(\{\omega \in \Omega : X(\omega) = x\})$$

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?

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

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

Linearity of Expectation

- Recall that $E[X] = \sum_{x \in \Omega'} x p_X(x)$.

Linearity of Expectation

- Recall that $E[X] = \sum_{x \in \Omega'} x p_X(x)$.
- Functions of random variables are random variables.

Linearity of Expectation

- Recall that $E[X] = \sum_{x \in \Omega'} x p_X(x)$.
- Functions of random variables are random variables.
- Furtermore, $E[g(X)] := \sum_{x \in \Omega'} g(x) P_X(x)$

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 $E[Y]$?

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 $E[Y]$?

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

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 $Y = aX + b$.

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 ?

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 $i \in \{1, 2, \ldots, n\}$.

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 $i \in \{1, 2, \ldots, n\}$.

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 $i \in \{1, 2, \ldots, n\}$.

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

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 ..

Towards Variance ..

- Recall $E[X] = \sum_{x \in \Omega'} x p_X(x)$.

Towards Variance ..

- Recall $E[X] = \sum_{x \in \Omega'} x p_X(x)$.
- Furthermore, $E[X^n] = \sum_{x \in \Omega'} x^n p_X(x)$.

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, $E[g(X)] := \sum_{x \in \Omega'} g(x) p_X(x)$

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 $g(X) = (X - E[X])^2$.

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.

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!

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.

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

Variance

- $E[g(X)] = E[(X - E[X])^2]$ is called as the variance of random variable X.

Variance

- $E[g(X)] = E[(X - E[X])^2]$ is called as the variance of random variable X.
- $Var(X) := E[(X - E[X])^2]$

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 $E[(X - E[X])^2] = E[X^2] - E[X]^2$

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.

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

- Indicator random variable $1_A(\omega) = \begin{cases} 1, & \text{If } \omega \in A \subseteq \Omega \\ 0, & \text{otherwise.} \end{cases}$

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 $p_{1_A}(x) = \begin{cases} \mathbb{P}(A), & \text{when } x = 1 \\ 1 - \mathbb{P}(A), & \text{when } x = 0. \end{cases}$

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.

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.

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 $E[1_A]$?

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

- Bernoulli random variable $X = \begin{cases} 1, & \text{with probability } p \\ 0, & \text{otherwise.} \end{cases}$

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.

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.

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.

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.

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.

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.

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

- Consider a biased coin (head with probability p) and toss it n times.

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.

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.

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?.

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?. $P_N(k) = \binom{n}{k} p^k (1-p)^{n-k}$.

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

Geometric random variable

- Consider a biased coin (head with probability p) and suppose you keep tossing it till head appears the first time.

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.

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?

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? $p_N(k) = (1-p)^{k-1}p$.

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

Poisson random variable

- A Poisson random variable X comes with a parameter λ and has $\Omega' = \mathbb{Z}_{\geq 0}$

Poisson random variable

- A Poisson random variable X comes with a parameter λ and has $\Omega' = \mathbb{Z}_{\geq 0}$
- PMF: $p_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}$

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.

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.

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.

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.