ECE 302: Chapter 03: Discrete Random Variables

Fall 2019

Prof Stanley Chan

School of Electrical and Computer Engineering
Purdue University
1. Random Variable
Definition

A random variable $X$ is a function $X : \Omega \rightarrow \mathbb{R}$ that maps an outcome $\xi \in \Omega$ to a number $X(\xi)$ on the real line.

Why need Random Variable?

- Coin flip: Head or Tail
- Alphabet: a, b, c, ..., z

We want to map these outcomes to numbers.
Example

- Flip a coin 2 times.
- The sample space $\Omega$ is
  \[ \Omega = \{(HH), (HT), (TH), (TT)\}. \]
- $\xi_1 = HH$, $\xi_2 = HT$, $\xi_3 = TH$, $\xi_4 = TT$.

Then,
- $X = \text{number of H}$.
- $X(\xi_1) = \# \text{ of heads in } \{HH\} = 2$
- $X(\xi_2) = \# \text{ of heads in } \{HT\} = 1$
- $X(\xi_3) = \# \text{ of heads in } \{TH\} = 1$
- $X(\xi_4) = \# \text{ of heads in } \{TT\} = 0$
Random Variable as a Mapping

Assign an outcome to a real number.
You can map multiple outcomes to the same real number.
What is the probability $X = a$?

Go back to the sample space, and find out $\xi_1$ and $\xi_2$

Measure the size of the event in the sample space
You want to calculate $P[X = a]$

So you find out

$P[\{\xi_1, \xi_2\}]$  \hspace{1cm} (1)

This is equivalent to

$P[\xi \mid \xi \in X^{-1}(a)]$  \hspace{1cm} (2)

$X^{-1}(a)$ is the inverse image of $a$ in the sample space.

$X^{-1}(a)$ in our example is $\{\xi_1, \xi_2\}$

$P$ can only measure things in $\Omega$. This is how probability is defined!

So you need to go back to the sample space and measure the set there.
Example 1

- Flip a coin 2 times.
- The sample space $\Omega$ is

$$\Omega = \{(\text{HH}), (\text{HT}), (\text{TH}), (\text{TT})\}.$$

- $\xi_1 = \text{HH}$, $\xi_2 = \text{HT}$, $\xi_3 = \text{TH}$, $\xi_4 = \text{TT}$.
- $X =$ number of H.

**What is the probability of getting 1 head?**

- What is the sample space?
- What is $X$? What is $a$?
- What is $X^{-1}(a)$?
- What are the elements we need to count?
Example 2

- Throw a dice 2 times.
- The sample space $\Omega$ is
  \[ \Omega = \{(1, 1), (1, 2), \ldots, (6, 6)\}. \]
- $\xi_1 = (1, 1), \xi_2 = (1, 2), \ldots, \xi_{36} = (6, 6)$.
- $X = \text{sum of two numbers}$.

**What is the probability of getting a 7?**

- What is the sample space?
- What is $X$? What is $a$?
- What is $X^{-1}(a)$?
- What are the elements we need to count?
Probability Mass Function

Definition

The **probability mass function (PMF)** of a random variable $X$ is a function which specifies the probability of obtaining a number $X(\xi) = a$. We denote a PMF as

$$p_X(a) = \mathbb{P}[X = a].$$

**Interpretation.** $\mathbb{P}[X = x]$ should be interpreted as

$$\mathbb{P}[X = a] = \mathbb{P}[\{\xi \mid X(\xi) = a\}] = \mathbb{P}[\{\xi \mid \xi \in X^{-1}(a)\}].$$
Difference between $X$ and $a$

The **random variable** $X$ ...
- is an object, not a numerical value!
- always has a PMF (for discrete case)
- has a different numerical value for a different experiment.

The **numerical value** $a$ ...
- is a number, a deterministic number
- is a particular outcome; we call it a random realization
- could a variable in an equation, e.g., $a^2 + 3a + 2 = 0$

When we write $P[X = a]$, we mean:

When we write $X = \text{numpy.random.rand}(1)$, we mean:
**Remark.** A PMF should satisfy the condition that

\[ \sum_{k \in X(\Omega)} p_X(k) = 1. \]
Examples

Example. Let $X$ be a random variable with PMF

$$p_X(k) = c \left( \frac{1}{2} \right)^k, \quad k = 1, 2, \ldots$$

Find $c$.

Example. Let $X$ be a random variable with PMF

$$p_X(k) = c \sin \left( \frac{\pi}{2} k \right),$$

for $k = 0, 1, 2, \ldots$ Find $c$. 
2. Expectation
Expectation

Definition (Expectation)

The **expectation** of a random variable \( X \) is

\[
E[X] = \sum_x x \, p_X(x).
\]

**Interpretation**: Weighted average.
True Mean and Sample Mean

**True Mean** \( \mathbb{E}[X] \)
- A statistical property of a random variable.
- A deterministic number.
- Often unknown, or is the center question of estimation.
- You have to know \( X \) in order to find \( \mathbb{E}[X] \); Top down.

**Sample Mean** \( \overline{X} \)
- A numerical value. Calculated from data.
- Itself is a random variable.
- It has uncertainty.
- Uncertainty reduces as more samples are used.
- We use sample mean to estimate the true mean.
- You do not need to know \( X \) in order to find \( \overline{X} \); Bottom up.
Histograms

$n = 100$

$n = 10000$
Computing the Expectation

Let $X$ be a random variable with PMF

$$p_X(k) = [0.3, c, 0.2, 0.4], \quad k = 1, 2, 3, 4$$

(a) Find $c$
(b) Find $E[X]$
Computing the Expectation

Let $X$ be a random variable with PMF

\[ p_X(k) = \frac{c}{2^k}, \quad k = 1, 2, \ldots \]

(a) Find $c$
(b) Find $\mathbb{E}[X]$
Example

Consider a game. Flip a coin 3 times. Reward:

- $1 if there are 2 Heads
- $8 if there are 3 Heads
- $0 if there are 0 or 1 Head

The cost to enter the game is $1.5. On average what is the net gain?
Properties of $\mathbb{E}[X]$

Property (1. Function of $X$)

For any function $g$, 

$$
\mathbb{E}[g(X)] = \sum_{x} g(x)p_X(x).
$$
Properties of \( \mathbb{E}[X] \)

Property (2. Linearity)

*For any function \( g \) and \( h \),*

\[
\mathbb{E}[g(X) + h(X)] = \mathbb{E}[g(X)] + \mathbb{E}[h(X)].
\]
Properties of $\mathbb{E}[X]$

Property (3. Scale)

*For any constant $c$,*

$$\mathbb{E}[cX] = c\mathbb{E}[X].$$
Properties of $\mathbb{E}[X]$

Property (4. DC Shift)

For any constant $c$,

$$\mathbb{E}[X + c] = \mathbb{E}[X] + c.$$
Moment

Definition

The \textit{k-th moment} of a random variable $X$ is

$$\mathbb{E}[X^k] = \sum_x x^k p_X(x).$$

Example. Flip a coin 3 times. Let $X$ be the number of heads. Then,

$$p_X(0) = \frac{1}{8}, \quad p_X(1) = \frac{3}{8}, \quad p_X(2) = \frac{3}{8}, \quad p_X(3) = \frac{1}{8}.$$ 

The second moment $\mathbb{E}[X^2]$ is
Variance

Definition

The variance of a random variable $X$ is

$$\text{Var}[X] = \mathbb{E}[(X - \mu_X)^2],$$

where $\mu_X = \mathbb{E}[X]$ is the expectation of $X$.

Example. $X =$ coin flip with probability $p$. Find variance of $X$. 
Properties of Variance

**Property**

*The variance of a random variable $X$ has the following properties*

(a) **Moment.**

$$\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$  

(b) **Scale.** *For any constant $c$,*

$$\text{Var}[cX] = c^2 \text{Var}[X].$$

(c) **DC Shift.** *For any constant $c,$*

$$\text{Var}[X + c] = \text{Var}[X].$$
3. Cumulative Distribution Function
Cumulative Distribution Function

Definition

The **cumulative distribution function** (CDF) of a discrete random variable $X$ is

$$F_X(x) \overset{\text{def}}{=} \mathbb{P}[X \leq x] = \sum_{x' \leq x} p_X(x').$$

**Interpretation:**

- CDF is the “integration” of PMF
- CDF is *well-defined* whereas PMF is not quite
- CDF works for both discrete and continuous random variables
Example. Consider a random variable $X$ with PMF

$$p_X(0) = \frac{1}{4}, \quad p_X(1) = \frac{1}{2}, \quad p_X(4) = \frac{1}{4}.$$ 

Find and sketch CDF.
Properties of CDF

1. The CDF is a sequence of increasing

2. \( F_X(+\infty) = \)

3. \( F_X(-\infty) = \)

4. At positions where \( p_X(x) > 0 \), there is always a
Properties of CDF

The height of each jump is

The solid dot is always on the
Question: How to generate random number from a PMF $p_X(k) = [0.1 \ 0.4 \ 0.2 \ 0.3]$?

Procedure:

1. Compute $F_X(k)$
2. Draw $U \sim \text{Uniform}(0, 1)$
3. Check which bin in $F_X(k)$ does $U$ fall into.
4. Common Discrete Random Variables
Common Discrete Random Variables

- **Bernoulli Random Variable**
  \[ p_X(1) = p, \quad p_X(0) = 1 - p. \]

- **Binomial Random Variable**
  \[ p_X(k) = \binom{n}{k} p^{n-k} (1 - p)^k \]

- **Geometric Random Variable**
  \[ p_X(k) = (1 - p)p^{k-1} \]

- **Poisson Random Variable**
  \[ p_X(k) = \frac{\lambda^k e^{-\lambda}}{k!} \]
4.1 Bernoulli Random Variable
Bernoulli Random Variable

Definition
Let $X$ be a **Bernoulli** random variable. Then, the PMF of $X$ is

$$p_X(0) = 1 - p, \quad p_X(1) = p,$$

where $0 < p < 1$ is called the Bernoulli parameter. We write

$$X \sim \text{Bernoulli}(p)$$

to say that $X$ is drawn from a Bernoulli distribution with a parameter $p$.

Example. Coin flip.
Bernoulli Example

Randomized Algorithm. Consider a large linear system:

\[
\begin{bmatrix}
y_1 \\
\vdots \\
y_N
\end{bmatrix}
= 
\begin{bmatrix}
a_{11} & a_{12} & \ldots & a_{1N} \\
\vdots & \vdots & \ddots & \vdots \\
a_{M1} & a_{M2} & \ldots & a_{MN}
\end{bmatrix}
\begin{bmatrix}
x_1 \\
\vdots \\
x_N
\end{bmatrix}
\]

where \( M \) and \( N \) are very large.

Example.

- Large network analysis (million-node network)
- Large-scale inverse problem (giga-pixel deconvolution)
- Large-scale routing (air traffic control)
- Large-scale decomposition (genome analysis)
Bernoulli Example

**Brute-force computation:**

\[ y_i = \sum_{j=1}^{N} a_{ij}x_j. \]

**Randomized computation:** Let \( l_j \sim \text{Bernoulli}(p_j) \). Then,

\[ \hat{y}_i = \sum_{j=1}^{N} a_{ij}x_j l_j / p_j, \]

We can prove that with extremely high probability, the deviation between \( \hat{y}_i \) and \( y_i \) is very small: As \( N \to \infty \),

\[ \mathbb{P}[|\hat{y}_i - y_i| > \epsilon] \to 0. \]

Bernoulli Example

Erdos-Renyi Graph

- A very famous (and simple) model for large networks.
- Is used nowadays to study network structures, e.g. Facebook, Google.

Graph

Matrix

4.2 Binomial Random Variable
Binomial Random Variable

Definition
Let $X$ be a **Binomial** random variable. Then, the PMF of $X$ is

$$p_X(k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \ldots, n,$$

where $0 < p < 1$ is the Binomial parameter, and $n$ is the total number of states. We write

$$X \sim \text{Binomial}(n, p)$$

to say that $X$ is drawn from a Binomial distribution with a parameter $p$ of size $n$.

**Interpretation**: Sum of $n$ Bernoulli random variables.

**Example**: Number of heads in $n$ coin flips.
Moments of Binomial

Property

If $X \sim \text{Binomial}(n, p)$, then

$$
\mathbb{E}[X] = np,
$$

$$
\mathbb{E}[X^2] = np(np + (1 - p)),
$$

$$
\text{Var}[X] = np(1 - p).
$$

Proof.
4.3 Geometric Random Variable
Geometric Random Variable

Definition
Let $X$ be a Geometric random variable. Then, the PMF of $X$ is

$$p_X(k) = (1 - p)^{k-1}p, \quad k = 1, 2, \ldots,$$

where $0 < p < 1$ is the Geometric parameter. We write

$$X \sim \text{Geometric}(p)$$

to say that $X$ is drawn from a Geometric distribution with a parameter $p$.

Interpretation: Flip a coin until you get a head.
Moments of Geometric Random Variables

**Property**

If $X \sim \text{Geometric}(p)$, then

\[
\begin{align*}
\mathbb{E}[X] &= \frac{1}{p}, \\
\mathbb{E}[X^2] &= \frac{2}{p^2} - \frac{1}{p}, \\
\text{Var}[X] &= \frac{1 - p}{p^2}.
\end{align*}
\]
4.4 Poisson Random Variable
Poisson Random Variable

Definition
Let \( X \) be a **Poisson** random variable. Then, the PMF of \( X \) is

\[
p_X(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 1, 2, \ldots ,
\]

where \( \lambda > 0 \) is the Poisson rate. We write

\( X \sim \text{Poisson}(\lambda) \)

to say that \( X \) is drawn from a Poisson distribution with a parameter \( \lambda \).

**Example.** Telephone arrivals. Photon arrivals. Passenger arrivals.
PMF of Poisson

\[ \text{PMF of Poisson} \]

\[ \lambda = 1 \]
\[ \lambda = 20 \]
\[ \lambda = 50 \]
Example 1

Photon Arrivals.

- Let $x$ be the intensity of a pixel. Assume $x \in [0, 1]$.
- Let $\lambda \overset{\text{def}}{=} \alpha x$ be the photon flux, i.e., number of photons per unit time. $\alpha > 1$ is a gain factor.

Let $Y$ be a random variable representing the number of photons. Then, the probability of getting $k$ photons is

$$P[Y = k] = \frac{\lambda^k e^{-\lambda}}{k!} = \frac{(\alpha x)^k e^{-\alpha x}}{k!}$$

- $\alpha = 10$
- $\alpha = 100$
- $\alpha = 1000$
Example 2

**Energy efficient escalators in airport.**
- These escalators have two modes: ON, or STAND-BY.
- When no pedestrian for $t_0$ seconds, turn to STAND-BY.
- On average people arrives at a rate of $\lambda$ people per second.
- How much energy is saved?

**Partial Solution.**
Let $N$ be the number of pedestrians for $t$ seconds.

$$\mathbb{P}[N = n] = \frac{(\lambda t)^n e^{-\lambda t}}{n!}.$$ 

Let $T$ be the inter-arrival time.

$$\mathbb{P}[T > t] = \mathbb{P}[\text{no arrival in } t] = \mathbb{P}[N = 0] = e^{-\lambda t}.$$ 

(Full answer in Ch. 4 when we introduce function of R.V.)
Moments of Poisson

Property

If $X \sim \text{Poisson}(\lambda)$, then

\[
\begin{align*}
\mathbb{E}[X] &= \lambda, \\
\mathbb{E}[X^2] &= \lambda + \lambda^2, \\
\text{Var}[X] &= \lambda.
\end{align*}
\]
**Poisson Binomial Approximation**

**Theorem (Poisson Approximation to Binomial)**

For small \( p \) and large \( n \), and let \( \lambda \overset{\text{def}}{=} np \),

\[
\binom{n}{k} p^k (1 - p)^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda}
\]
Example

Problem.
- Data arrival rate: $n = 10^9$ bits per second.
- Probability of having one error bit: $p = 10^{-9}$.
- In one second, how likely will we get $k = 5$ error bits?

Solution.
If you stick to binomial:
- Binomial: Flip coin $10^9$ times. Get 5 heads.
  $\binom{10^9}{5} (10^{-9})^5 (1 - 10^{-9})^{10^9 - 5}$.
- Numerical problem!!

If we use Poisson approximation:
- $\lambda = np = (10^9)(10^{-9}) = 1$.
  $\frac{1^5}{5!} e^{-1}$.
- Much easier!
Proof (not required)

Let \( \lambda = np \). Then,

\[
\binom{n}{k} p^k (1 - p)^{n-k}
\]

\[
= \frac{n!}{k!(n-k)!} \left( \frac{\lambda}{n} \right)^k \left( 1 - \frac{\lambda}{n} \right)^{n-k}
\]

\[
= \frac{\lambda^k}{k!} \frac{n(n-1) \ldots (n-k+1)}{n \cdot n \cdot \ldots \cdot n} \left( 1 - \frac{\lambda}{n} \right)^{n-k}
\]

\[
= \frac{\lambda^k}{k!} \left( \frac{1}{n} \right) \left( \frac{1}{n} \right) \ldots \left( \frac{1}{n} \right) \left( 1 - \frac{\lambda}{n} \right)^{n-k} \left( 1 - \frac{\lambda}{n} \right)^n
\]

\[
\rightarrow 1 \text{ as } n \rightarrow \infty
\]

\[
= \frac{\lambda^k}{k!} \left( 1 - \frac{\lambda}{n} \right)^n.
\]
Proof (not required)

We claim that \((1 - \frac{\lambda}{n})^n \rightarrow e^{-\lambda}\). This can be proved by noting that

\[
\log(1 + x) \approx x, \quad x \ll 1.
\]

It then follows that \(\log (1 - \frac{\lambda}{n}) \approx -\frac{\lambda}{n}\). Hence, \((1 - \frac{\lambda}{n})^n \approx e^{-\lambda}\)