# Introduction to probability - MA2110

Amit Tripathi

Indian Institute of Technology, Hyderabad

August 9, 2019

# Linearity of expectation

**Lemma**

$E[X + Y] = E[X] + E[Y]$

# Meaning of expectation

What do we mean by expectation?

# A notation

Expectation $E[X]$ is also called mean and often denoted as $\mu_X$ or just $\mu$ (when underlying random variable is clear).

# Example

## Example

Let $X$ be a random variable with probability mass function given as

$$P(X = -1) = .2 \quad P(X = 0) = .5 \quad P(X = 1) = .3$$

Compute $E[X^2]$.
We define a new random variable $Y = X^2$. Note that $Y$ can take values 0 and 1. We find the probability mass function $Y$ as follows:

$$P(Y = 0) = 0.5 \quad P(Y = 1) = .5$$

Therefore

$$E[X^2] = E[Y] = 0 \times P(Y = 0) + 1 \times P(Y = 1) = 0.5$$

# Theorem

**Theorem**

*[Law of the unconscious student]* *Let $X$ be a discrete random variable with probability mass function $MF_X(x)$. Assume that for some function $G$, $G(X)$ also defines a random variable. Then*

$$E[G(X)] = \sum_{x\,:\,MF(x)>0} G(x)\, MF_X(x)\, dx$$

# Indicator random variable

## Definition

Let $A$ be any event. We define the **indicator random variable $I_A$** of $A$ as

$$I_A = \begin{cases} 1, & A \text{ happens} \\ 0, & A \text{ doesn't happens} \end{cases}$$

# Variance

## Definition

Let $X$ be a RV with mean $\mu$. Then the **variance** of $X$, denoted $Var(X)$ is defined by

$$Var(X) = E[(X - \mu)^2]$$

Usually denoted as $\sigma_X^2$.

# Variance

**Lemma**

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

**Proof.**

We have

$$Var(X) = E[(X - \mu)^2] = E[X^2] - 2\mu E[X] + \mu^2 = E[X^2] - \mu^2$$

$\square$

# Variance

Let $A$ be an event with $P(A) = p$ and let $I_A$ denotes the indicator random variable. Find $Var(I_A)$ ?

# Standard deviation

**Definition**

The square root of $Var(X)$ is called the **standard deviation of** $X$ and usually denoted as $\sigma_X$.

**Tip:** When in doubt whether to use $\sigma_X$ or $\sigma_X^2$ remember that $\sigma_X$ has same unit as mean $E[X]$.

# Question

Suppose $a$ and $b$ are constants. What is $Var(aX + b)$ ?

# Bernoulli random variables

**Definition**

An independent trial is performed with probability of success $= p$
Let $X$ denotes the number of success. Then $X$ is said to be a **Bernoulli random variable** with parameter $p$.

The probability mass function of a Bernoulli RV with parameter $p$ is given by

$$P(X = 0) = 1 - p, \quad \text{and } P(X = 1) = p$$

# Bernoulli random variables - expectation

Let $X = Bernoulli(p)$. Then

$$\begin{aligned} E[X] =& 0 \cdot P(X = 0) + 1 \cdot P(X = 1) \\ =& 0 + 1 \cdot p \\ =& p \end{aligned}$$

Similarly

$$\begin{aligned} Var(X) =& E[(X - E[X])^2] \\ =& (0 - p)^2 \, P(X = 0) + (1 - p)^2 \, P(X = 1) \\ =& p^2(1 - p) + (1 - p)^2 p \\ =& p(1 - p) \end{aligned}$$

## Binomial random variable

Fix $0 \leq p \leq 1$ and $n \in \mathbb{N}$.

### Definition

If $n$ independent trials are performed, each of which may result in a success with probability $p$ and failure with probability $(1-p)$. Let $X$ denotes the number of success. Then $X$ is said to be a **Binomial random variable** with parameters $(n, p)$.

# Binomial random variable

The probability mass function of a binomial RV with parameters $(n, p)$ is given by

$$P(X = r) = MF_X(r) = \begin{cases} \binom{n}{r} p^r (1-p)^{n-r} & r \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

**Question**

If $X = Bin(n, p)$, then what is the value of

$$P(X = 0) + P(X = 1) + \cdots + P(X = n) + \cdots$$

**Obvious point** - Binomial with parameters $(n, p)$ is $n$ independent trials of Bernoulli with parameter $p$.

# Binomial random variable - expectation

**Lemma**

*Let $X$ be a binomial RV with parameters $(n, p)$. Then*

1. $E[X] = np.$
2. $Var(X) = np(1-p)$

# Binomial random variable - expectation

**Expectation.**

Let $I_k$ be the indicator random variable that the $k$'th toss is a head. We know that $E[I_k] = p$. Then

$$X = I_1 + I_2 + \cdots + I_n$$

Therefore

$$E[X] = np$$

□

# Question

A coin with probability of head $= p$ is tossed $n$ times. Let $H$ denotes the number of heads and $T$ denotes the number tails. Find $E[H - T]$ ?

# A brief distraction

Let $X$ and $Y$ be two discrete random variables.

**Question**

What should we mean by $P(X = a, Y = b)$ ?

**Definition**

We say $X$ and $Y$ are <u>**independent random variables**</u> if

$$P(X = a, Y = b) = P(X = a)P(Y = b)$$

for all $a, b$.

Suppose $X = Bin(m, p)$ and $Y = Bin(n, p)$ be two independent binomial random variable.

**What is the distribution of** $X + Y$ **?**

# Binomial random variable - variance

**Fact: (to be proved later)** If $I_1, \cdots, I_n$ are independent random variables then

$$Var(I_1 + I_2 + \cdots + I_n) = Var(I_1) + Var(I_2) + \cdots + Var(I_n)$$

Using this, what is $Var(X)$ when $X = Bin(n, p)$ ?

# Binomial random variables

**Lemma**

*Let $X$ be a binomial RV with parameters $(n, p)$. Then its probability mass function*

$$p(r) = \binom{n}{r} p^r (1-p)^{n-r}$$

*satisfies following:*

1. $p(r) > p(r-1) \iff r < (n+1)p$
2. $p(r) < p(r-1) \iff r > (n+1)p$
3. $p(r) = p(r-1) \iff r = (n+1)p$

# Example of Binomial distribution

## Example

Three candidates in an election. The numbers of voters is $2N$ and each votes for

1. candidate 1 with probability $p_1$,
2. candidate 2 with probability $p_2$ and
3. candidate 3 with probability $p_3$.

A candidate wins the election if he/she secures at least half votes.
**Question: Find probability that candidate 1 will win.**

Every voter has two choices - vote for candidate 1 with probability $p_1$ or doesn't votes for that candidate with probability $p_2 + p_3 = 1 - p_1$. Then candidate 1 will win the election with probability

$$P_1 = \sum_{r=N}^{2N} \binom{2N}{r} p_1^r q_1^{2N-r}.$$

# Geometric random variable

Independent trials, each having a probability $0 < p < 1$ are performed until a success occurs. Suppose $X$ denote she number of trials required then the probability mass function is given as

$$P(X = n) = (1 - p)^{n-1} p$$

**Exercise**

Verify that $\sum_{i=1}^{\infty} P(X = i) = 1$

## Example

An urn contains $w$ white and $b$ black balls. Balls are randomly selected, one at a time, until a black ball is obtained. ASSUME that each ball selected is replaced before the next one is drawn. Find the probability that

1. Exactly $n$ draws are needed. Note that this is a case of geometric probability with $p = \dfrac{b}{a+b}$. Thus we have $P(X = n) = \dfrac{a^{n-1}b}{(a+b)^n}$.
2. At least $k$ draws are needed. We have $P(X \geq k) = \dfrac{a^{k-1}}{(a+b)^{k-1}}$ as the first $k-1$ draws should be failure!

# Expected value

Let $X$ be a geometric random variable with PMF
$P(X = n) = (1-p)^{n-1}p$. Find $E[X]$ ?
**Can you guess it intuitively?**

$$
\begin{aligned}
E[X] &= \sum_{i=1}^{\infty} i(1-p)^{i-1}p \\
&= \sum_{i=1}^{\infty} (i-1)(1-p)^{i-1}p + \sum_{i=1}^{\infty} (1-p)^{i-1}p \\
&= \sum_{j=0}^{\infty} j(1-p)^{j}p + 1 \\
&= (1-p)E[X] + 1
\end{aligned}
$$

Therefore $E[X] = \dfrac{1}{p}$.

# Expectation of non-negative integer valued random variables

**Lemma**

*Suppose $X$ be a non-negative integer valued random variable. Then*

$$E[X] = \sum_{i=0}^{\infty} P(X > i)$$

If $X$ is geometric, we can apply the above result.

# Expected value of a geometric random variable (Again)

Suppose $X$ is geometric - then by Lemma

$$E[X] = \sum_{i=0}^{\infty} P(X > i)$$

We know that $P(X > i)$ is $(1-p)^i$. Thus we get

$$E[X] = \sum_{i=0}^{\infty} (1-p)^i = \frac{1}{1-(1-p)} = \frac{1}{p}$$

# Variance

Let $X$ be a geometric random variable with PMF

$$P(X = n) = (1-p)^{n-1}p$$

Then variance of $X$ is

$$E[X^2] = \sum_{i=1}^{\infty} i^2(1-p)^{i-1}p$$

which simplifies to $E[X^2] = \dfrac{2-p}{p^2}$. Therefore

$$Var(X) = E[X^2] - E[X]^2 = \frac{1-p}{p^2}$$

# Probability on natural numbers

**Question**

Can you describe a probability measure on the set of natural numbers ?

One possibility $P(X = k) = \frac{1}{2^k}, k = 1, 2, 3, \cdots.$

# Density of subsets of natural numbers

Suppose $A \subset \mathbb{N}$. We define **density of** $A$ as

$$P(A) = \lim_{n \to \infty} \frac{|A \cap \{1, 2, 3, \cdots, n\}|}{n}$$

1. $P(A) = \dfrac{1}{2}$ when $A$ is the set of even numbers.
2. $P(A) = \dfrac{1}{7}$ when $A$ is the set of multiples of 7.
3. $P(A) = \dfrac{6}{\pi^2}$ when $A$ is the set of square free numbers.

**Question**- Does $P$ defines a probability on $\mathbb{N}$?

# Coupon collector problem

**Question**

Consider a child collecting different kinds of superhero coupons. There are $N$ kinds of superheroes and hence so many different coupons. **Assume that any coupon is equally likely to be any of the superhero.** How many coupons the child should expect to buy before he/she completes the collection of all $N$ superheroes.

# Applications

Applications in engineering, statistics and even biology!

# Case $N = 3$

$X =$ (random) number of coupons that we need to purchase to complete our collection.

This is a random variable and we want to find $E[X]$.

1. Let $X_1$ denotes the number of coupons we need to purchase to have 1 superman coupons. What is $X_1$?
2. Let $X_2$ denotes the additional number of coupons that we need to purchase a **new** superman coupon (we already have **one** superman coupon). What is $X_2$ in the previous examples?
3. Let $X_3$ denotes the additional number of coupons that we need to purchase another **new** superman coupons (we already have **two** superman coupons). What is $X_3$ in the previous examples?

# Case $N = 3$

Clearly $X = X_1 + X_2 + X_3$. In particular

$$E[X] = E[X_1] + E[X_2] + E[X_3].$$

1. What is $E[X_1]$ ?
2. What is $E[X_2]$ ?
3. What is $E[X_3]$ ?

Therefore

$$E[X] = 3\left(\frac{1}{1} + \frac{1}{2} + \frac{1}{3}\right).$$

# General case

More generally, we have $N$ different coupons and

$$X = X_1 + X_2 + \cdots + X_N.$$

Arguing as before

$$E[X_i] = \frac{N}{N - i + 1}$$

Thus

$$E[X] = N\left(\frac{1}{1} + \frac{1}{2} + \cdots \frac{1}{N}\right).$$

# Two questions

### Question

Suppose you are given a random number generator which generates a number between $(0, 1)$ everytime randomly. Can you use it to simulate a coin toss with probability of success $p$ ?

### Question

Suppose you are given a coin toss with probability of success $p$. Can you use it to simulate a fair coin toss?

# Poisson distribution

**Definition**

A random variable is said to satisfy **Poisson distribution** if the probability mass function of $X$ satisfies $p(i) = e^{\lambda}\dfrac{\lambda^i}{i!}$ for $i = 0, 1, 2, \dots$

Poisson distribution is indeed a distribution

$$\sum_{i=0}^{\infty} p(i) = \sum_{i} \exp^{-\lambda} \frac{\lambda^i}{i!} = \exp^{-\lambda} \exp^{\lambda} = 1$$

# Poisson distribution - expected value

$$
\begin{aligned}
E[X] &= \sum_{k=0}^{\infty} \frac{ke^{-\lambda}\lambda^k}{k!} \\
&= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \\
&= \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \quad \text{where } j = k-1 \\
&= \lambda
\end{aligned}
$$

# Poisson distribution - variance

$$
\begin{aligned}
E[X^2] &= \sum_{k=0}^{\infty} \frac{k^2 e^{-\lambda}\lambda^k}{k!} \\
&= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{k\lambda^{k-1}}{(k-1)!} \\
&= \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{(j+1)\lambda^j}{j!} \quad \text{where } j = k-1 \\
&= \lambda \left[ \sum_{j=0}^{\infty} \frac{je^{-\lambda}\lambda^j}{j!} + \sum_{j=0}^{\infty} \frac{e^{-\lambda}\lambda^j}{j!} \right] \\
&= \lambda(\lambda+1)
\end{aligned}
$$

Since $E[X] = \lambda$, we get

$$Var(X) = E[X^2] - (E[X])^2 = \lambda$$

# Poisson distribution

$$X \sim Poi(\lambda).$$

# Poisson distribution - the law of rare events

**Rare events can be modeled using Poisson distribution:**

- **Number of bankruptcies filed in a month.**
- **Number of large meteorites striking earth over a year.**
- **Number of accidents on a road stretch over a time period.**
- **Number of high magnitude earthquakes (say 7 or higher) occurring in a region over a year.**
- **Number of times certain extreme weather condition occurs over a time period.**
- **Radioactive decay.**

# Poisson distribution - when does it apply

Our interest is in finding the number of times some event occur in a given interval (of time, length etc). We assume that these events satisfy the following:

1. We can divide the given interval into smaller subintervals such that the probability of more than one success in any subinterval is zero;
2. the probability of one success in a subinterval is proportional to its length;
3. Subintervals are independent of each other.

Then we can apply Poissonian distribution.

# Poisson distribution - quick questions

**Question**

Do the assumptions above imply $P(X > 1) = 0$ ?

**Question**

Suppose $X \sim Poi(\lambda_1)$ and $Y \sim Poi(\lambda_2)$. Assume that $X$ and $Y$ are independent random variables. Let $Z = X + Y$. Find

1. $P(Z = 0)$ ?
2. $P(Z = 1)$ ?
3. $P(Z = 2)$ ?

# Poisson Paradigm

Suppose $\{E_i\}_{i=1}^{n}$ be a collection of events satisfying

1. $P(E_i) = p_i \ll 1$.
2. Further suppose that the events $E_i$ are weakly dependent, that is, $P(E_i|E_j) \sim P(E_i)$.
3. Assume that the number of trials $n$ is large enough.

Let $X$ denotes the number of events $E_i$ that occur in $n$ trials, then

$$P(X = k) \simeq \exp(-\lambda)\frac{\lambda^k}{k!}$$

where $\lambda = p_1 + p_2 + \cdots + p_n$ and $k$ is small.

# Prussian Army and Horse Kick Deaths

Between 1875 to 1894, it was found that 122 men died due to horse kicks among **ten Prussian army corps**.

1. In most years in most corps, no one died from being kicked.
2. But in one corp in one year, four men were kicked to death.

**Question**

Does this mean something was amiss in this particular corp?

# Prussian Army and Horse Kick Deaths

Ladislaus Bortkiewicz - a Russian statistician studied this problem. Was it an issue with that corp or was this expected?

**Total corp years =** $20 \times 10 = 200$**. Total deaths =** $122$**.**

Table: Data of Horse Kick Deaths

| Deaths/corp year | Corp Years | Deaths |
|---:|---:|---:|
| 0 | 109 | 109 x 0 = 0 |
| 1 | 65 | 65 x 1 = 65 |
| 2 | 22 | 22 x 2 = 44 |
| 3 | 3 | 3 x 3 = 9 |
| 4 | 1 | 4 x 1 = 4 |
| 5+ | 0 | 0 |

Mean $\lambda = 122/200 = 0.610$ deaths per corp year. Let $X$ be the number of deaths in a corp year.

$$X \sim Poi(\lambda)$$

# Prussian Army and Horse Kick Deaths

Table: Data of Horse Kick Deaths

| Deaths/corp year | Corp Years | Deaths | Prob | Predicted |
|---:|---:|---:|---:|---:|
| 0 | 109 | 109 x 0 = 0 | 0.54335 | 108.7 |
| 1 | 65 | 65 x 1 = 65 | 0.33144 | 66.3 |
| 2 | 22 | 22 x 2 = 44 | 0.10109 | 20.2 |
| 3 | 3 | 3x3 = 9 | 0.02056 | 4.1 |
| 4 | 1 | 4x1 = 4 | 0.00313 | 0.6 |
| 5+ | 0 | 0 | 0.00042 | .08 |

So we expect $.00313 \times 200 = 0.6$ times that the event "4 deaths in a corp year" will happen.

# Approximating terms of Binomial distribution in 19th century

Historically - Poisson distribution was introduced as an approximation to Binomial distribution.

**Suppose $X = Bin(n, p)$ and let $\lambda = np$.**

Assume that $n, p, k$ satisfy the following assumptions:

1. $n$ is large, $p$ is sufficiently small such that $np$ is of moderate size.
2. $n \gg k$.

# Approximating terms of Binomial distribution in 19th century

With above assumptions, let $\lambda = np$. Mass function of a binomial random variable $X = Bin(n, p)$ is

$$P(X = k) = \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k}$$

We look at this expression term by term

$$\begin{aligned}
\frac{n!}{(n-k)!k!} p^k &= \frac{n!}{(n-k)!k!} \left(\frac{\lambda}{n}\right)^k \quad \text{using } p = \lambda/n \\
&= \frac{n(n-1)\cdots(n-k+1)}{n^k} \frac{\lambda^k}{k!} \\
&\sim \frac{\lambda^k}{k!} \quad \text{using } n \gg k
\end{aligned}$$

# Approximating terms of Binomial distribution in 19th century

$$(1-p)^{n-k} = \left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$= \frac{\left(1-\frac{\lambda}{n}\right)^{n}}{\left(1-\frac{\lambda}{n}\right)^{k}}$$

For $n$ large compared to $k$, and $\lambda$ moderate (compared to $n$), we have

$$\left(1-\frac{\lambda}{n}\right)^{n} \simeq \exp(-\lambda) \;\; and \;\; \left(1-\frac{\lambda}{n}\right)^{k} \simeq 1$$

# Approximating terms of Binomial distribution in 19th century

Summing up

$$\frac{n!}{(n-k)!k!}p^k \to \frac{\lambda^k}{k!}$$

and

$$(1-p)^{n-k} \to \exp(-\lambda)$$

**Theorem**

*(De-Moivre) For $n \gg k$ and $\lambda = np$ of moderate size,*

$$\binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} \to \exp(-\lambda)\frac{\lambda^k}{k!}$$

# Approximating terms of Binomial distribution in 18th century

**Remark**

This find is not just useful theoretically but was also remarkably useful for 18th century mathematicians as there were no computers and calculating binomial coefficients is very computationally intensive task.

# Continuous random variables

**Definition**

A random variable $X$ is said to be **continuous** if there exists a function $f$ satisfying the following

1. $f(x) \geq 0$ for all $x$.
2. $\int_{-\infty}^{\infty} f(x)dx = 1$.
3. For any $a \leq b$, we have

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

**Definition**

Such a function $f$ as above is called the probability density function (PDF) of $X$.

# Probability density function

**Question**

For a continuous random variable $X$, what is $P(X = a)$ ?

**Caution:** Density function $f(x)$ itself doesn't represents probability of anything!

# Relation between cumulative distribution function and probability density function

By definition, cumulative distribution function

$$F(a) := P(X \leq a)$$

Thus if $X$ has pdf $f(x)$ then

$$F(a) := P(X \leq a) = \int_{-\infty}^{a} f(x)dx$$

Differentiating both sides w.r.t $a$ yields

$$\frac{d}{da}F(a) = f(a)$$

# An example

## Example

The time in hours that a electronic equipment functions before breaking down is a continuous random variable with PDF given by

$$f(x) = \begin{cases} \lambda\, exp^{-x/100} & x \geq 0 \\ 0 \;\; x < 0 \end{cases}$$

Find the probability that

1. Equipement will function between 50 to 100 hours before breaking down.
2. It will function for fewer than 100 hours.

First we compute $\lambda$ using $\int_{-\infty}^{\infty} f(x)\, dx = P(-\infty < x < \infty) = 1$. This gives

$$\lambda = \frac{1}{100}$$