Introduction to probability

Amit Tripathi

Indian Institute of Technology, Hyderabad

August 29, 2019

Independent random variables

Definition

We say that a finite sequence of random variables $\{X_1, X_2, \cdots, X_n\}$ is **independent** if for every $a_1, \cdots, a_n \in \mathbb{R}$

$$P(X_1 \leq a_1, X_2 \leq a_2, \cdots, X_n \leq a_n) = \prod_{k=1}^{n} P(X_k \leq a_k)$$

Pairwise independence doesn't imply independence

Example

Let X, Y be outcome of two fair coins. Let $Z = X \oplus Y$ (where $\oplus$ denotes exclusive OR). Clearly the joint distribution of X, Y, Z is given as

$$P(X = 0, Y = 0, Z = 0) = \frac{1}{4}$$

$$P(X = 0, Y = 1, Z = 1) = \frac{1}{4}$$

$$P(X = 1, Y = 0, Z = 1) = \frac{1}{4}$$

$$P(X = 1, Y = 1, Z = 0) = \frac{1}{4}$$

Pairwise independence doesn't imply independence

Example

One can easily check that pairwise $\{X, Y\}$, $\{X, Z\}$ and $\{Y, Z\}$ are independent. For instance

$$
\begin{aligned}
P(X = 0, Z = 1) =& P(X = 0, Y = 1, Z = 1) \\
=& \frac{1}{4} = P(X = 0)P(Z = 1)
\end{aligned}
$$

But it is clear that X, Y, Z are not independent as Z is determined by X, Y. For example

$$P(X = 0, Y = 0, Z = 0) = \frac{1}{4} \neq \frac{1}{8} = P(X = 0)P(Y = 0)P(Z = 0)$$

Independent random variables

We say that a sequence of random variables $\{X_1, X_2, \cdots,\}$ is independent if **every finite subsequence is independent.**

Identically distributed random variables

We say that a sequence of random variables $\{X_1, X_2, \cdots ,\}$ is **identically distributed** if they all have same probability distribution i.e. for all $a \in \mathbb{R}$,

$$F_{X_i}(a) = F_{X_j}(a), \quad \forall \; i, j$$

If a sequence of random variables $\{X_1, X_2, \cdots ,\}$ is **i**ndependent and **i**dentically **d**istributed, we say it is a sequence of **i.i.d** random variables.

Recall the following facts about variance:

1. $Var(a \cdot X + b) = a^2 \cdot Var(X)$.

2. If $X_1, \cdots, X_n$ are independent then

$$Var(X_1 + X_2 + \cdots + X_n) = Var(X_1) + Var(X_2) + \cdots + Var(X_n)$$

Central limit theorem

Let $X_1, \cdots, X_n, \cdots$ be a sequence of i.i.d random variables each having mean μ and variance σ^2 (both assumed to be finite). Define sample mean

$$S_n = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

Question: What is $E[S_n]$? Answer: μ.

Question: What is $Var(S_n)$?

Central limit theorem

Lemma

Let $X_1, X_2, \cdots, X_n$ be independent and identically distributed random variables having expected value μ and variance σ^2 then variance

$$Var(S_n) = \frac{\sigma^2}{n}$$

Proof.

$$Var(S_n) = Var\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{1}{n^2} Var(\sum_{i=1}^n X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}. \qquad \square$$

Central limit theorem

Question: What is the expected value of $\dfrac{S_n - \mu}{\sigma/\sqrt{n}}$?

Question: What is the variance of $\dfrac{S_n - \mu}{\sigma/\sqrt{n}}$?

Theorem (CLT)

If $\{X_1, \cdots\}$ be an i.i.d sequence then the distribution function of the normalized variable $\dfrac{S_n - \mu}{\sigma/\sqrt{n}}$ approaches the standard normal distribution, i.e.

$$\lim_{n\to\infty} P\left\{\frac{S_n - \mu}{\sigma/\sqrt{n}} \leq a\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} \exp(-x^2/2)\, dx$$

Central limit theorem

Alternatively, it can be written in terms of the sum

$$X_1 + X_2 + \cdots + X_n = n \cdot S_n$$

Theorem (Central Limit Theorem)

With notation as above,

$$\lim_{n\to\infty} P\left\{\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leq a\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} \exp(-x^2/2)\, dx$$

Theorem (Central Limit Theorem)

With notation as above,

$$\lim_{n\to\infty} P\left\{\frac{S_n - \mu}{\sigma/\sqrt{n}} \leq a\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} \exp(-x^2/2)\, dx$$

Central limit theorem

There are various versions of central limit theorem. Most of them replace the conditions "identically distributed" and "independent distribution" by some weaker or different condition.

CLT is applicable to the case of discrete random variables as well.

Covariance of two random variable, provides us with a numerical measure of how they vary jointly.

Definition

We define **covariance of X and Y** as

$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$$

Question: What is $Cov(X,X)$?

Covariance

By definition

$$Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

This simplifies to

$$E[(X - \mu_X)(Y - \mu_Y)] = E[XY - \mu_X \cdot Y - \mu_Y \cdot X + \mu_X \cdot \mu_Y]$$

$$= E[XY] - \mu_X \mu_Y$$

Covariance

Question: Suppose X and Y are independent, what is the $Cov(X, Y)$?

Conversely, suppose $Cov(X, Y) = 0$, can we expect X and Y to be independent?

Covariance

Example

Consider a discrete random variable X with

$$P(X=-1) = P(X=0) = P(X=1) = \frac{1}{3}$$

Let $I_{X=0}$ be an indicator random variable defined as

$$I_{X=0} = \begin{cases} 1, & \text{if } X = 0 \\ 0, & \text{if } X \neq 0 \end{cases}$$

Question: What is $X \cdot I_{X=0}$? Thus $E[X \cdot I_{X=0}] = 0$.

Question: What is $Cov(X, I_{X=0})$? On the other hand,

$$P(X=1, I_{X=0}=0) = P(X=1) = \frac{1}{3} \neq P(X=1)P(I_{X=0}=0)$$

Covariance

To sum up:

$$X, Y \text{ are independent} \implies Cov(X, Y) = 0$$

But

$$Cov(X, Y) = 0 \not\Longrightarrow X, Y \text{ are independent}$$

Properties of covariance

$$Cov(X, Y) = E[XY] - \mu_X \mu_Y.$$

1. $Cov(X, Y) = Cov(Y, X)$.
2. $Cov(X_1 + X_2, Y) = Cov(X_1, Y) + Cov(X_2, Y)$.
3. $Cov(aX, Y) = aCov(X, Y)$.

Properties of covariance

$$Cov(X, Y) = E[XY] - \mu_X \mu_Y.$$

1. $Cov\left(X_1 + X_2, Y_1 + Y_2\right) =$
   $Cov(X_1, Y_1) + Cov(X_1, Y_2) + Cov(X_2, Y_1) + Cov(X_2, Y_2).$

2. More generally,
   $Cov\left(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j\right) = \sum_i \sum_j Cov(X_i, Y_j).$

Covariance

$$Cov(X, Y) = E[XY] - \mu_X \mu_Y.$$

1. $Cov\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_i \sum_j Cov(X_i, Y_j).$

Corollary

$Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i) + \sum_{i=1}^{n} \sum_{j \neq i} Cov(X_i, X_j)$

Proof.

Use the fact that $Var(\sum_{i=1}^{n} X_i) = Cov(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i)$. □

Covariance

Question

Suppose $\{X_i\}$ are independent random variables. Show that $Var(\sum_{i=1}^n X_i) = \sum_{i=1}^n Var(X_i)$.

Proof.

Previous result shows

$$Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i) + \sum\sum_{i \neq j} Cov(X_i, X_j)$$

and when X_i are independent then $Cov(X_i, X_j) = 0$. $\square$

Covariance

Question

Let X, Y have equal variance. Find $Cov(X+Y, X-Y)$?

Proof.

$$\begin{aligned} Cov(X+Y, X-Y) =& Cov(X,X) + Cov(Y,X) \\ &+ Cov(X,-Y) + Cov(Y,-Y) \end{aligned}$$

$Cov(Y,X) + Cov(X,-Y) = Cov(X,Y) - Cov(X,Y) = 0.$
Therefore

$$Cov(X+Y, X-Y) = Cov(X,X) - Cov(Y,Y) = 0$$

$\square$

Covariance

Question: If $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ be two **independent normal random variables** then what is the variance of $X + Y$? It is given as

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2$$

Correlation coefficient

Definition

The correlation coefficient of X and Y is defined to be

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}$$

provided that $Var(X)$ and $Var(Y)$ are both positive.

Correlation coefficient

Lemma

$-1 \leq \rho(X, Y) \leq 1.$

Proof.

Let $Var(X) = \sigma_X^2 > 0$ and $Var(Y) = \sigma_Y^2 > 0$.

$$\begin{aligned} 0 \leq & Var\left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right) \\ = & Var\left(\frac{X}{\sigma_X}\right) + Var\left(\frac{Y}{\sigma_Y}\right) - 2 \cdot \frac{Cov(X, Y)}{\sigma_X \sigma_Y} \end{aligned}$$

$\square$

Correlation coefficient

Proof contd.

$$\begin{aligned} 0 \leq & Var\left(\frac{X}{\sigma_X}\right) + Var\left(\frac{Y}{\sigma_Y}\right) - 2 \cdot \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \\ = & 2(1 - \rho(X,Y)) \end{aligned}$$

Therefore $\rho(X,Y) \leq 1$. A similar argument shows the lower bound. $\square$

Correlation coefficient

It is clear from the proof that $\rho(X, Y) = 1$ if and only if $Var\left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right) = 0$.

Fact: If $Var(Z) = 0$ then $P(Z = E[Z]) = 1$.

Thus if $\rho(X, Y) = 1$ then with probability 1,

$$\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y} = K$$

for some constant K. Rewriting

$$Y = aX + b$$

for some constants a, b where $a > 0$.

To sum up

Theorem

Let X, Y be two random variables.

1. *If $\rho(X, Y) = 1$ then with probability 1,*

$$Y = aX + b$$

   for some constants a, b where $a > 0$.

2. *If $\rho(X, Y) = -1$ then with probability 1,*

$$Y = aX + b$$

   for some constants a, b where $a < 0$.

Correlation coefficient

In general, a high correlation **only suggests but DOES NOT IMPLY** some relationship between the random variables.

Example

If A, B, C, D are pairwise uncorrelated random variables, each with mean 0 and variance 1. Find

1. $\rho(A+B, C+D)$.
2. $\rho(A+B, B+C)$.

Conditional expectation

Suppose we want to know the **conditional probability** of $X = x$ **given that** $Y = y$. These are two events and hence the corresponding probability is given as

$$P(X = x | Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$

where we are (as in the case of conditional probability) forced to assume that $P(Y = y) > 0$.

Conditional expectation

Thus we define

Definition

The **conditional probability mass function** of X, **given that** $Y = y$, is defined, for all y such that $P(Y = y) > 0$, by

$$p_{X|Y}(x|y) = P(X = x|Y = y) = \frac{p(x, y)}{p_Y(y)}$$

Conditional expectation

Question

Is this really a probability mass function?

$$p_{X|Y}(x|y) = P(X = x|Y = y) = \frac{p(x,y)}{p_Y(y)}$$

Conditional expectation

Once we have above definition, the following doesn't appear too surprising

Definition

The **conditional expectation of** X **given that** $Y = y$, for all values of y such that $p_Y(y) > 0$ is defined as

$$E[X|Y = y] = \sum_x x\, P(X = x|Y = y) = \sum_x x p_{X|Y}(x|y)$$

Conditional expectation

Example

If $Y \sim U(0,1)$ and

$$P(X = x | Y = y) = \begin{cases} \binom{n}{x} y^x (1-y)^{n-x} & \text{for } x \in \{0, 1, \cdots, n\} \\ 0 & \text{otherwise} \end{cases}$$

Find $E[X \mid Y = y]$.

Conditional density function

Similarly, for continuous case, we define

Definition

If $f(x, y)$ be the joint density function of X and Y then we define the **conditional density function** of X **given that** $Y = y$ and $f_Y(y) > 0$, as

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

Conditional expectation

Lemma

Conditional density function is a density function i.e for any y

$$\int f_{X|Y}(x|y)dx = 1$$

Proof.

$$\int f_{X|Y}(x|y)dx = \int \frac{f_{X,Y}(x,y)}{f_Y(y)}dx = \frac{\int f_{X,Y}(x,y)dx}{f_Y(y)} = \frac{f_Y(y)}{f_Y(y)} = 1$$

$\square$

Conditional expectation

Definition

We define **conditional expectation** of X, given $Y = y$, as

$$E[X|Y = y] = \int_{-\infty}^{\infty} x\, f_{X|Y}(x|y)\, dx$$

provided $f_Y(y) > 0$.

Conditional expectation

Example

Suppose the joint density function of X and Y is given by

$$f(x,y) = \begin{cases} \dfrac{e^{-x/y}e^{-y}}{y}, & 0 < x, y < \infty \\ 0, & \text{otherwise} \end{cases}$$

Find $E[X|Y = y]$?

Conditional expectation

Example

We first calculate the conditional probability density of X as

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{\left(\dfrac{e^{-x/y}e^{-y}}{y}\right)}{\left(\displaystyle\int_0^\infty \frac{e^{-x/y}e^{-y}}{y}dx\right)}$$

$$= \frac{\left(\dfrac{e^{-x/y}e^{-y}}{y}\right)}{e^{-y}} = \frac{e^{-x/y}}{y}$$

Thus

$$E[X|Y=y] = \int_0^\infty x\frac{e^{-x/y}}{y}dx = y$$

Conditional expectation

Question: - We know that $E[X|Y = y]$ is a number. How to interpret $E[X|Y]$?

We think of it as a function $g(Y)$ such that the value at $Y = y$ is

$$g(y) = E[X|Y = y]$$

Conditional expectation

Question: - What about $E[E[X|Y]]$? Is this a number or a random variable?

It turns out it is an interesting number but first an example.

Conditional expectation

Example

Let X be a discrete random variable and let I_A be indicator random variable, corresponding to some event A.

Question: Find $E[E[X|I_A]]$?

By definition $E[g(Y)] = \sum_y g(y)P(Y = y)$ and $I_A \in \{0, 1\}$.
Therefore

$$E[E[X|I_A]] = E[X|I_A = 0]P(I_A = 0) + E[X|I_A = 1]P(I_A = 1)$$

Conditional expectation

Example

$$E[E[X|I_A]] = E[X|I_A = 0]P(I_A = 0) + E[X|I_A = 1]P(I_A = 1)$$

$$E[X|I_A = 0]P(I_A = 0) = \sum_x \frac{xP(X = x, I_A = 0)}{P(I_A = 0)} P(I_A = 0)$$

$$E[X|I_A = 1]P(I_A = 1) = \sum_x \frac{xP(X = x, I_A = 1)}{P(I_A = 1)} P(I_A = 1)$$

Therefore

$$\begin{aligned} E[E[X|I_A]] &= \sum_x x\left(P(X = x, I_A = 0) + P(X = x, I_A = 1)\right) \\ &= \sum_x xP(X = x) = E[X] \end{aligned}$$

Conditional expectation

Proposition

$E[X] = E[E[X|Y]]$.

Proof.

We take the continuous case:

$$
\begin{aligned}
E[E[X|Y]] &= \int E[X|Y=y] f_Y(y) dy \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx \right) f_Y(y) dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \left(\frac{f_{X,Y}(x,y)}{f_Y(y)} f_Y(y) \right) dx\, dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x\, f_{X,Y}(x,y) dx\, dy \\
&= E[X]
\end{aligned}
$$

Markov's inequality

Proposition

Let X be a random variable that takes only **non-negative** *values. Then for any $\alpha > 0$ we have*

$$P(X \geq \alpha) \leq E[X]/\alpha$$

[Observe that only two assumptions are made - first is $X \geq 0$ and second is a consequence of the first as otherwise $P(X \geq \alpha) = 1$ and this inequality is trivially false!]

Markov's inequality

Proof.

Define an indicator random variable

$$I_X = \begin{cases} 1 & \text{if } X \geq \alpha \\ 0 & \text{otherwise} \end{cases}$$

Since $X \geq 0$ and $\alpha > 0$, we always have

$$I_X \leq \frac{X}{\alpha}$$

In particular

$$E[I_X] \leq \frac{E[X]}{\alpha}$$

But

$$E[I_X] = P(X \geq \alpha) \cdot 1 + P(X < \alpha) \cdot 0 = P(X \geq \alpha)$$

Markov's inequality

Note that when $\alpha < E[X]$ then Markov's inequality is trivially true. Also for known distributions, one can directly compute and obtain a better bound. Importance of this result is **usually** evident when

1. α is large compared to $E[X]$ and
2. We don't know anything about the distribution of a RV except the expected value.

Markov's inequality

Example

Let $X \sim Bin(n,p)$. Then we know that $E[X] = np$. Then for any $a \in \mathbb{N}$,

$$P(X \geq a) = 1 - P(X < a) = 1 - \sum_{k=0}^{a-1} \binom{n}{k} p^k (1-p)^{n-k}$$

whereas by Markov's inequality, we are getting

$$P(X \geq a) \leq \frac{E[X]}{a} = \frac{np}{a}$$

One can check that bound obtained directly is better than Markov's inequality.

Markov's inequality

Example

Average annual rainfall in India is 300 mm. What can be said about the probability that in 2019 it will exceed 500 mm?

By Markov's inequality we have

$$P(X \geq 500) \leq E[X]/500 = 3/5$$

Note that as α becomes larger compared to $E[X]$ the bound becomes smaller - thus e.g.

$$P(X \geq 1500) \leq 1/5$$

Chebyshev's inequality

Proposition

Let X be a random variable with **finite mean** μ *and variance* σ^2. *Then for any* $\alpha > 0$

$$P(|X - \mu| \geq \alpha) \leq \sigma^2/\alpha^2$$

Proof.

Applying Markov's inequality for the random variable $(X-\mu)^2$ and α^2, (and using the fact that by definition $E[(X-\mu)^2] = \sigma^2$) we obtain

$$P((X-\mu)^2 \geq \alpha^2) \leq \sigma^2/\alpha^2$$

This is equivalent to our claim. $\square$

Chebyshev's inequality

Corollary (Alternate formulation of Chebyshev's inequality)

Putting $k = \alpha/\sigma$ we get

$$P(|X - \mu| \geq k\sigma) \leq 1/k^2$$

Chebyshev's inequality

1. As before, for known distributions, one can always obtain better bounds than this. But when nothing is known except expected value and variance then this bound can be very useful.
2. It is easy construct example of a random variable for which the bound is strict - Take X to be a discrete RV with $P(X = a) = 1/2$ and $P(X = -a) = 1/2$. Then $E[X] = 0$ and $\sigma^2 = Var(X) = E[X^2] = a^2$. Thus

$$P(|X - \mu| \geq a) = P(|X| \geq a) = 1 = \sigma^2/a^2$$

Chebyshev's inequality

Example

There was a numerical error in the version done in the class. Edited version has a more realistic value of variance. Average annual rainfall in India is 300 mm. **Suppose in addition the variance is known to be 400mm square .** What can be said about the probability that in 2019 it will exceed 500 mm?

By Markov's inequality - $P(X \geq 500) \leq E[X]/500 = 3/5$

Question: How to obtain Chebychev's inequality in this question?

$$\begin{aligned} P(X \geq 500) =& P(|X - 300| \geq 200) - P(X \leq 100) \\ \leq& \frac{\sigma^2}{a^2} - P(X \leq 100) \\ =& \frac{400}{200 * 200} - P(X \leq 100) \\ \leq& \frac{1}{100}, \text{ This improves Markov's bound sharply!} \end{aligned}$$

The Weak law of large numbers

Assuming that $\{X_i\}$ have independent and identical distribution with mean μ and variance $= \sigma^2$.

As before, define **sample mean** as

$$S_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Then

$$Var(S_n) = \sigma^2/n$$

The Weak law of large numbers

Weak law of large numbers is a trivial corollary to Chebyshev's inequality:

Theorem (The weak law)

Let $X_1, X_2, \cdots,$ be a sequence of independent and identically distributed random variables with finite mean $E[X_i] = \mu$. Then for any $\delta > 0$

$$\lim_{n\to\infty} P\left\{|S_n - \mu| \geq \delta\right\} = 0$$

The Weak law of large numbers

Proof.

Recall that $Var(S_n) = \sigma^2/n$.

Applying Chebyshev's inequality it follows that

$$P\left\{|S_n - \mu| \geq \delta\right\} \leq \frac{\sigma^2}{n\delta^2}$$

Therefore

$$\lim_{n\to\infty} P\left\{|S_n - \mu| \geq \delta\right\} = 0$$

$\square$

The Weak law of large numbers

Theorem (The weak law - alternative description)

With notation as above, for any $\delta > 0$

$$\lim_{n\to\infty} P\left\{|S_n - \mu| \leq \delta\right\} = 1$$

Discrete random variables

Example

We toss a coin with $P(H) = p$. Find the probability mass function of expected waiting time for getting total of k heads.

Suppose W_k denotes the number of toss needed to get k heads. This means that the last toss was a head and out of previous $n-1$ tosses, there were $k-1$ heads thus

$$P(W_k = n) = \binom{n-1}{k-1} p^{k-1} \cdot q^{n-k} \cdot p = \binom{n-1}{k-1} p^k q^{n-k}$$

Binomial distribution

Example

Fix some $n \in \mathbb{N}$ and $0 \leq p \leq 1$. We consider n-noded **simple** graphs. What is simple graph?

Index the nodes as $1, 2, \cdots, n$ and construct a random graph by connecting the nodes i and i with probability p. The degree of vertex i, designated as D_i, is the number of edges that have vertex i as one of their vertices.

1. Find the probability mass function of D_i.
2. Find the correlation coefficient $\rho(D_i, D_j)$.

Binomial distribution

Example

1. $P(D_i = k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}$ Thus it is a Binomial distribution with parameters $(n-1, p)$.
2. First we want $Var(D_i)$. What is it? Since $D_i \sim Bin(n-1, p)$ therefore $Var(D_i) = (n-1)p(1-p)$.

   If $i = j$ then clearly $\rho(D_i, D_j)$ is equal to 1.

   When $i \neq j$, are D_i and D_j independent?

Binomial distribution

Example

We recall that if $E =$ number of all the edges in the graph then

$$2E = \sum_i D_i$$

What is the probability mass function of E?

$$P(E = k) = \binom{N}{k} p^k (1-p)^{N-k}, \quad \text{where } N = \frac{n(n-1)}{2}$$

Therefore $Var(E) = Np(1-p) = \dfrac{n(n-1)}{2} p(1-p)$.

Binomial distribution

Example

$$Var(E) = Np(1-p) = \frac{n(n-1)}{2}p(1-p).$$

We assume that $Cov(D_i, D_j) = \lambda$ whenever $i \neq j$ then

$$Var(2E) = Cov(\sum_i D_i, \sum_j D_j) = \sum_i Var(D_i) + n(n-1)\lambda$$

Therefore

$$Cov(D_i, D_j) = p(1-p)$$

This means

$$\rho(D_i, D_j) = \frac{Cov(D_i, D_j)}{\sqrt{Var(D_i)Var(D_j)}} = \frac{p(1-p)}{(n-1)p(1-p)} = \frac{1}{n-1}$$

Continuous random variable

Question

A certain type of disease affects $\alpha\%$ of the population annually. Find the probability that out of N people randomly selected, not more than m will have the disease next year (assume $m \ll N$ and $\alpha \ll 100\%$).

Underlying assumption - any person having this disease is an independent event irrespective of who else has this disease.

Continuous random variable

Question

Any given person has this disease with probability $p = \alpha/100$. Clearly this is a Binomial distribution with parameters (N, p). Therefore the probability is

$$\sum_{k=0}^{m} \binom{N}{k} p^k (1-p)^{N-k}$$

Theorem

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

$$\binom{n}{k} p^k (1-p)^{n-k} \to \frac{e^{-\lambda} \cdot \lambda^k}{k!}$$

Continuous random variable

Question

With our assumptions, we can approximate this using Poisson distribution with parameter

$$\lambda = Np$$

Therefore the probability is

$$\sum_{k=0}^{m} \binom{N}{k} p^k q^{N-k} = \sum_{k=0}^{m} \frac{e^{-\lambda}\lambda^k}{k!}$$

Maximum random variable

Definition

Given random variables $X_1, \cdots, X_n$. Define

$$X = \max(X_1, \cdots, X_n).$$

Question: Is $X = X_i$ for some i ?

Question: Let $\{X_i\}$ be independent random variables and denote the distribution function of X_i by $F_{X_i}(x)$. Find the distribution function of X.

Maximum random variable

Solution

We have

$$\begin{aligned} F(a) =& P(Z \le a) \\ =& P(\max(X_1, X_2, \cdots, X_n) \le a) \\ =& P(X_1 \le a, X_2 \le a, \cdots, X_n \le a) \end{aligned}$$

Applying independence

$$F(a) = P(X_1 \le a)P(X_2 \le a) \cdots P(X_n \le a) = F_1(a)F_2(a) \cdots F_n(a)$$

Joint density function

Suppose the joint density function of random variables X and Y is given as

$$f(x,y) = \begin{cases} \lambda x e^{-x} & 0 < x < \infty, 0 < y < 2 \\ 0 & \text{otherwise} \end{cases}$$

Find the value of λ.

Joint density function

Integral of joint distribution function over $\mathbb{R}^2$ plane must be 1. Therefore

$$\int_0^\infty \int_0^2 \lambda x e^{-x} \, dx \, dy = \int_0^\infty 2 \cdot \lambda x e^{-x} \, dx = 1$$

which gives $\lambda = \dfrac{1}{2}$.

Joint density function

The joint density function of X and Y is

$$f(x,y) = \begin{cases} ax + by, & 0 < x < 1, 0 < y < 1 \\ 0, & \text{otherwise} \end{cases}$$

Are X and Y independent?

Joint density function

Question

Two random variables are independent if and only if their joint probability density (mass) function can be expressed as

$$f_{X,Y}(x,y) = h(x)g(y) \qquad -\infty < x < \infty, -\infty < y < \infty$$

One side is trivial. Which one?

Conversely, suppose

$$f_{X,Y}(x,y) = h(x)g(y) \qquad -\infty < x < \infty, -\infty < y < \infty$$

Joint density function

Question

Therefore we get

$$f_X(x) = \int f_{X,Y}(x, y)\, dy = \int h(x)g(y)\, dy = \lambda h(x)$$

for some constant λ. Similarly

$$f_Y(y) = \int f_{X,Y}(x, y)\, dx = \int h(x)g(y)\, dy = \mu g(y)$$

for some constant μ.

What is the value of $\lambda \cdot \mu$?

Joint density function

Question

We know that

$$\int f_X(x)\,dx = \lambda \int h(x)\,dx = \lambda \cdot \mu$$

Therefore $\lambda \cdot \mu = 1$. This proves that

$$f_{X,Y}(x,y) = \lambda \cdot \mu\, h(x)g(y) = f_X(x) \cdot f_Y(y)$$

Therefore X and Y are independent.

Sum of random variables

Question

Suppose X_i, $i = 1, 2, \cdots, 50$ are independent random variables with $X_i \sim Poi(\lambda_i)$ for positive constants λ_i, $i = 1, 2, \cdots, 50$. Let

$$X = X_1 + X_2 + \cdots X_{50}$$

Find the distribution of X ?

We look at the case of only 2 Poisson random variable $X = Poi(\lambda_1)$ and $Y = Poi(\lambda_2)$.

Sum of random variables

Question

The probability mass function of $X + Y$ is

$$\begin{aligned}
P(X + Y = n) =& \sum_{k=0}^{n} P(X = k, Y = n - k) \\
=& \sum_{k=0}^{n} P(X = k)P(Y = n - k) \\
=& \sum_{k=0}^{n} e(-\lambda_1)\frac{\lambda_1^k}{k!} e(-\lambda_2)\frac{\lambda_2^{n-k}}{(n-k)!} \\
=& \frac{e(-\lambda_1 - \lambda_2)\sum_{k=0}^{n}\binom{n}{k}\lambda_1^k\lambda_2^{n-k}}{n!} \\
=& e(-\lambda)\frac{\lambda^n}{n!}
\end{aligned}$$

Sum of random variables

Question

Thus $X + Y$ is again a poisson random variable with parameter $\lambda = \lambda_1 + \lambda_2$. Now use induction to get the general case!

Question

Let X, Y be independent exponential random variables with parameters λ_1, λ_2 respectively.

1. Find distribution of $Z = X/Y$?
2. Compute $P(X < Y)$?

Recall that exponential density function is given as

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

Exponential random variables

Question

Let $Z = X/Y$. Then

$$P(Z \le a) = P(X \le aY)$$

Definition

We will say that X and Y are jointly continuous if there exists a function $f(x,y)$ such that for every set $C \subset \mathbb{R}^2$ we have

$$P(\{X,Y\} \in C) = \iint_{(x,y)\in C} f(x,y)\,dx\,dy$$

Question: What is the joint density function of X, Y ?

Exponential random variables

Question

$$
\begin{aligned}
P(Z \le a) =& P(X \le aY) \\
=& \int_0^\infty \int_0^{ay} \lambda_1 \lambda_2 e^{-\lambda_1 x - \lambda_2 y} \, dx \, dy \\
=& \lambda_2 \int_0^\infty e^{-\lambda_2 y} \left(1 - e^{-a\lambda_1 y}\right) \, dy \\
=& 1 - \frac{\lambda_2}{a\lambda_1 + \lambda_2} \\
=& \frac{a\lambda_1}{\lambda_2 + a\lambda_1}
\end{aligned}
$$

1. Find distribution of $Z = X/Y$? $P(Z \le a) = \dfrac{a\lambda_1}{\lambda_2 + a\lambda_1}$.
2. Compute $P(X < Y)$? Putting $a = 1$ we get answer to the second question as $\dfrac{\lambda_1}{\lambda_1 + \lambda_2}$.

Central limit theorem

Question

Suppose X_i, $i = 1, 2, \cdots, 50$ are independent random variables with $X_i \sim Poi(\lambda)$ where $\lambda = 0.03$. Let $X = X_1 + X_2 + \cdots X_{50}$.

1. Use CLT to evaluate $P(X \geq 3)$.

Can we apply CLT in this case?

Question

Suppose X_i, $i = 1, 2, \cdots, 50$ are independent random variables with $X_i \sim Poi(\lambda)$ where $\lambda = 0.03$. Let $X = X_1 + X_2 + \cdots X_{50}$.

1. What is $E[X]$? $E[X] = 1.5$.
2. What is $Var[X]$? $Var[X] = 50 \cdot Var(X_1) = 50 \cdot 0.03 = 1.5$.

Therefore

$$\begin{aligned} P(X \geq 3) =& P\left(\frac{X - E[X]}{\sigma_X} \geq \frac{3 - E[X]}{\sigma_X}\right) \\ =& P\left(\frac{X - E[X]}{\sigma_X} \geq \frac{3 - 1.5}{\sqrt{1.5}}\right) \\ =& 1 - \Phi(\sqrt{1.5}) \end{aligned}$$

where $\Phi(x) = P(Y \leq x)$ for $Y \sim \mathcal{N}(0, 1)$.

Notation

In the exam, you may use the symbol

$$\Phi(x) = P(Y \leq x) := \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

i.e. $Y \sim \mathcal{N}(0, 1)$.

But make sure to mention that $\Phi(x)$ denotes the CDF of standard normal distribution!

Airline ticket

Example

We toss a fair coin 1000 times. Find the smallest k such that we can say with probability 95% that number of heads is less than or equal to k.

We apply CLT - here $X_i \sim Ber\left(\frac{1}{2}\right)$ and $\mu = \frac{1}{2}$ and $\sigma^2 = \frac{1}{4}$. We are interested in smallest k such that

$$P\left(\sum_i X_i \leq k\right) = 0.95$$

Applying CLT

$$P\left(\frac{\sum_i X_i - 1000 \cdot \mu}{\sigma\sqrt{1000}} \leq \frac{k - 1000 \cdot \mu}{\sigma\sqrt{1000}}\right) \simeq \Phi\left(\frac{k - 1000 \cdot \mu}{\sigma\sqrt{1000}}\right)$$

Example

From tables of standard normal distribution, we get $\Phi(1.645) \simeq 0.95$. Therefore

$$\left(\frac{k - 1000 \cdot \mu}{\sigma\sqrt{1000}}\right) = 1.645$$

Or

$$k = 1000 \cdot \mu + 1.645 \times \sigma \times \sqrt{1000} \simeq 500 + 26 = 526$$

Conditional expectation

Example

Suppose $X \in Bin(n, p)$ & $Y \in Bin(m, p)$. Find $E[X \,|\, X + Y = k]$?

Definition

The **conditional expectation of** X **given that** $Z = z$, for all values of z such that $p_Z(z) > 0$ is defined as

$$E[X|Z = z] = \sum_x x\, P(X = x|Z = z) = \sum_x x\, p_{X|Z}(x|z)$$

Conditional expectation

Example

Suppose $X \in Bin(n,p)$ & $Y \in Bin(m,p)$ be independent random variables. Find $E[X \,|\, X+Y=k]$?

Question: What is the distribution of $X+Y$? Answer: $X+Y \sim Bin(m+n,p)$.

We calculate the conditional probability mass function of X as

$$
\begin{aligned}
P(X=j|X+Y=k) &= \frac{P(X=j, Y=k-j)}{P(X+Y=k)} \\
&= \frac{\binom{n}{j} p^j q^{n-j} \binom{m}{k-j} p^{k-j} q^{m-k+j}}{\binom{m+n}{k} p^k q^{m+n-k}} \\
&= \frac{\binom{n}{j}\binom{m}{k-j}}{\binom{m+n}{k}}
\end{aligned}
$$

Conditional expectation

Example

Thus we get

$$
\begin{aligned}
E(X|X+Y=k) &= \sum_{j=0}^{n} \frac{\binom{n}{j}\binom{m}{k-j}\cdot j}{\binom{m+n}{k}} \\
&= \sum_{j=1}^{n} \frac{\binom{n}{j}\binom{m}{k-j}\cdot j}{\binom{m+n}{k}} \\
&= n\sum_{j=1}^{n} \frac{\binom{n-1}{j-1}\binom{m}{k-j}}{\binom{m+n}{k}} \\
&= \frac{n\cdot\binom{m+n-1}{k-1}}{\binom{m+n}{k}} \\
&= \frac{n\cdot k}{m+n}
\end{aligned}
$$

Conditional expectation

Example

A miner is trapped in a mine containing 3 doors.

1. The first door will lead him to safety in 3 hrs.
2. The second door will take him to a tunnel that will return him back to the mine in 5 hrs.
3. The third door will take him to a tunnel that will return him back to the mine in 7 hrs.

If the miner is equally likely to choose any one of the doors, what is the expected length of time until he reaches safety.

Conditional expectation

Example

Let X denotes the amount of time for the miner to reach some safe point. Let Y be the door that he initial chooses. Then

$$\begin{aligned} E[X] =& E[E[X|Y]] \\ =& E[X|Y=1]P(Y=1) + E[X|Y=2]P(Y=2) \\ & + E[X|Y=3]P(Y=3) \\ =& \frac{1}{3}\left(E[X|Y=1] + E[X|Y=2] + E[X|Y=3]\right) \end{aligned}$$

Question: What are the values of $E[X|Y=1]$, $E[X|Y=2]$ and $E[X|Y=2]$?

Conditional expectation

Example

Then

$$E[X] = \frac{1}{3}(3 + 5 + E[X] + 7 + E[X])$$

Solving this gives $E[X] = 15$.

Table of well known distributions

Distributions	Binomial	Geometric	Poisson	Uniform$[a,b]$	Exponential	Normal
Mass function	$\binom{n}{r}p^r q^{n-r}$	$q^{n-1}p$	$e^{-\lambda}\frac{\lambda^i}{i!}, i \geq 0$			
Density function				$\frac{1}{b-a}$	$\lambda e^{-\lambda x},\ x \geq 0$	$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
$E[X]$	np	$\frac{1}{p}$	λ	$\frac{a+b}{2}$	$\frac{1}{\lambda}$	μ
$Var(X)$	npq	$\frac{q}{p^2}$	λ	$\frac{(b-a)^2}{12}$	$\frac{1}{\lambda^2}$	σ^2