Probability & Statistics (1)

Random Variables (II)

Asst. Prof. Chan, Chun-Hsiang

Master program in Intelligent Computing and Big Data, Chung Yuan Christian University, Taoyuan, Taiwan
Undergraduate program in Intelligent Computing and Big Data, Chung Yuan Christian University, Taoyuan, Taiwan
Undergraduate program in Applied Artificial Intelligence, Chung Yuan Christian University, Taoyuan, Taiwan

Outlines

1. Random Variables
2. Discrete Random Variables
3. Expected Value
4. Expectation of a Function of a Random Variable
5. Variance
6. The Bernoulli and Binomial Random Variables
7. The Poisson Random Variable
8. Other Discrete Probability Distributions
9. Expected Value of Sums of Random Variables
10. Properties of the Cumulative Distribution Function
11. [#9] Assignment
12. Reference
13. Question Time

The Poisson Random Variable

- 在生活中，Poisson random variable/ distribution的應用領域其實相當的廣泛:
  1. 在一本書中，每一頁(章節)中會有多少的錯字
  2. 一天中，打錯多少次電話
  3. 每小時，可以等到多少班的公車
  4. 每一天，會有多少的客人來到商店
  5. **(舉個例子。。。)**
  6. 。。。
  7. 。。。

The Poisson Random Variable

- **Poisson random variable (卜瓦松分配)**

A random variable X that takes on one of the values $0,1,2,\ldots$ is said to be a Poisson random variable with parameter λ if, for some $\lambda > 0$,

$$p(i) = P\{X = i\} = e^{-\lambda}\frac{\lambda^i}{i!}, where\ i = 0,1,2,\ldots$$

The abovementioned defines a probability mass function, since

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

The Poisson Random Variable

The Poisson random variable could be approximated for a binomial random variable with parameters (n, p) when n is large and p is small enough so that np is of moderate size.

Suppose that X is a binomial random variable with parameters (n, p) and let $\lambda = np$. Then

$$P\{X = i\} = \frac{n!}{(n-i)!\,i!} p^i (1-p)^{n-i} = \frac{n!}{(n-i)!\,i!} \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i}$$

$$= \frac{n(n-1)\cdots(n-i+1)}{n^i} \frac{\lambda^i}{i!} \frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^i}$$

The Poisson Random Variable

For n large and λ moderate,

$$\left(1-\frac{\lambda}{n}\right)^n \approx e^{-\lambda}$$

$$\frac{n(n-1)\cdots(n-i+1)}{n^i} \approx 1$$

$$\left(1-\frac{\lambda}{n}\right)^i \approx 1$$

Therefore,

$$P\{X=i\} \approx e^{-\lambda}\frac{\lambda^i}{i!}$$

The Poisson Random Variable

The expected value and variance of Poisson random variable

已知binomial random variable的期望值與變異數分別為np與$np(1-p)$。在計算Poisson random variable的期望值與變異數之前，我們知道當n很大且p很小的時候，我們可以用binomial random variable來做近似。因此期望值就可以用$\lambda = np$；同理可證，變異數就可以得出$\lambda(1-p) \approx \lambda$。

有發現到一個很有趣的現象嗎?

Poisson random variable的期望值與變異數剛好都是λ。

The Poisson Random Variable

- **Proof of expected value of Poisson random variable**

$$E[X] = \sum_{i=0}^{\infty} \frac{ie^{-\lambda}\lambda^i}{i!} = \lambda \sum_{i=1}^{\infty} \frac{e^{-\lambda}\lambda^{i-1}}{(i-1)!}$$

$$since\ j = i - 1$$

$$= \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!}$$

$$since\ \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = e^{\lambda}$$

$$= \lambda$$

The Poisson Random Variable

- **Proof of variance of Poisson random variable**

According to the definition of variance, $Var[X] = E[X^2] - [E[X]]^2$

$$E[X^2] = \sum_{i=0}^{\infty} \frac{i^2 e^{-\lambda} \lambda^i}{i!} = \lambda \sum_{i=1}^{\infty} \frac{i e^{-\lambda} \lambda^{i-1}}{(i-1)!}$$

$let\ j = i - 1$

$$= \lambda \sum_{j=0}^{\infty} \frac{(j+1) e^{-\lambda} \lambda^j}{j!} = \lambda \left[\sum_{j=0}^{\infty} \frac{j e^{-\lambda} \lambda^j}{j!} + \sum_{j=0}^{\infty} \frac{e^{-\lambda} \lambda^j}{j!} \right] = \lambda(\lambda + 1)$$

$$\Rightarrow Var[X] = E[X^2] - [E[X]]^2 = \lambda(\lambda + 1) - \lambda^2 = \lambda$$

The Poisson Random Variable

- **範例一**

今天你在晶圓廠做品管師，假設一張晶圓上面壞損的晶片數為 Poisson distribution ($\lambda = \frac{1}{2}$)，試問一張晶圓上至少有一個晶片是壞掉的機率為何?

Solution:

令X為一張晶圓上壞掉的晶片數量，則

$$P\{X \geq 1\} = 1 - P\{X = 0\} = 1 - e^{-\frac{1}{2}} \frac{\left(\frac{1}{2}\right)^0}{0!} \approx 0.393$$

The Poisson Random Variable

- 範例二

假設你今天在電子元件廠工作，某台機器所生產的元件壞掉率為0.1。試問隨機取樣10個元件中，最多一個壞掉的機率為何?

Solution:

Binomial approach:

$$\binom{10}{0}(0.1)^0(0.9)^{10}+\binom{10}{1}(0.1)^1(0.9)^9=0.7361$$

Poisson approach: $\lambda=np\Rightarrow 10\times 0.1=1$

$$P\{X=0\}+P\{X=1\}=e^{-1}\frac{1^0}{0!}+e^{-1}\frac{1^1}{1!}=e^{-1}+e^{-1}\approx 0.7358$$

The Poisson Random Variable

- **範例三**

假設你今天做粒子實驗，每一秒鐘一克的放射線物質可以釋放出數顆α粒子。根據過去經驗數據可以得知，平均大概會釋放出3.2顆α粒子。試問不釋放超過2顆α粒子的機率為何?

Solution:

$$\lambda = 3.2 \Rightarrow P\{X = i\} = e^{-3.2}\frac{3.2^i}{i!}$$

$$P\{X \le 2\} = e^{-3.2}\frac{3.2^0}{0!} + e^{-3.2}\frac{3.2^1}{1!} + e^{-3.2}\frac{3.2^2}{2!} \approx 0.3799$$

The Poisson Random Variable

Let the number of changes that occur in a given continuous interval be conducted. Then we have an appropriate Poisson process with parameter $\lambda > 0$, if the following conditions are satisfied.

1. The number of changes occurring in non-overlapping interval are independent.
2. The probability of exactly one change in a sufficiently short interval of length h is approximated to λh, where $h \to 0$.
3. The probability of two or more changes occurring a sufficiently short interval is zero.

The Poisson Random Variable

- **範例四**

假設在台灣南部地區地震在一周內發生次數符合Poisson的三個假設且$\lambda = 2$，試問:

(1) 再接下來兩周，至少三次地震發生的機率為何?

(2) 求從現在開始到下一次地震發生的機率分布(probability distribution)。

The Poisson Random Variable

Solution:

(a)

在前方的推導中，我們得知$P\{N(t)=k\}=e^{-\lambda t}\frac{(\lambda t)^k}{k!}, where\ k=0,1,\ldots$

$$P\{N(2)\geq 3\}=1-P\{N(2)=0\}-P\{N(2)\leq 1\}-P\{N(2)\leq 2\}$$

$$=1-e^{-2\times 2}\frac{(2\times 2)^0}{0!}-e^{-2\times 2}\frac{(2\times 2)^1}{1!}-e^{-2\times 2}\frac{(2\times 2)^2}{2!}$$

$$=1-e^{-4}-4e^{-4}-\frac{4^2}{2}e^{-4}=1-13e^{-4}$$

The Poisson Random Variable

(b)

令X為到下一次地震前的週數，因為X大於t，當今天如果在下個t時間區間內發生地震，

$$P\{X > t\} = P\{N(t) = 0\} = e^{-\lambda t}$$

$$F(t) = P\{X \le t\} = 1 - P\{X > t\} = 1 - e^{-\lambda t} = 1 - e^{-2t}$$

Computing the Poisson Distribution Function

- If X is Poisson with parameter λ, then

$$\frac{P\{X=i+1\}}{P\{X=i\}}=\frac{\frac{e^{-\lambda}\lambda^{i+1}}{(i+1)!}}{\frac{e^{-\lambda}\lambda^{i}}{i!}}=\frac{\lambda}{i+1}$$

Starting with $P\{X=0\}=e^{-\lambda}$,

$$P\{X=1\}=\lambda P\{X=0\}$$

$$P\{X=2\}=\frac{\lambda}{2}P\{X=1\}$$

$$\vdots$$

$$P\{X=i+1\}=\frac{\lambda}{i+1}P\{X=i\}$$

Computing the Poisson Distribution Function

- **[加分題] 範例五**

利用Python撰寫一個遞迴函數來計算以下兩題:

(a) 試求出$P\{X \leq 90\}$，當X為Poisson random variable且平均值為100。

(b) 試求出$P\{X \leq 1075\}$，當X為Poisson random variable且平均值為1000。

Ans:

(a) 0.1714; (b) 0.9894

Other Discrete Probability Distributions
* The Geometric Random Variable

- **The Geometric Random Variable (幾何分配)**

假設在數個獨立試驗中，成功的機率為$0 < p < 1$，試驗持續進行一直到成功發生為止。令X為需要的試驗數量，

$$P\{X = n\} = (1-p)^{n-1}p, where\ n = 1,2,\ldots$$

到第n次才成功，代表前面第1次到第$n-1$次都是失敗。

$$\sum_{n=1}^{\infty} P\{X = n\} = p\sum_{n=1}^{\infty}(1-p)^{n-1} = \frac{p}{1-(1-p)}, since\ S_\infty = \frac{p}{1-r}$$

$$= \frac{p}{p} = 1$$

Other Discrete Probability Distributions * The Geometric Random Variable

- **Proof of expected value of geometric random variable**

$$E[X] = \sum_{i=1}^{\infty} iq^{n-1}p = \sum_{i=1}^{\infty} (i-1+1)q^{i-1}p$$

$$= \sum_{i=1}^{\infty} (i-1)q^{i-1}p + \sum_{i=1}^{\infty} q^{i-1}p$$

$$= \sum_{j=0}^{\infty} jq^{j}p + 1, since\ j = i-1 \Rightarrow q\sum_{j=0}^{\infty} jq^{j-1}p + 1 = qE[X] + 1$$

$$since\ pE[X] = 1, E[X] = \frac{1}{p}$$

Other Discrete Probability Distributions * The Geometric Random Variable

- **Proof of variance of geometric random variable**

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

$$= \sum_{i=1}^{\infty} (i-1)^2 q^{i-1} p + \sum_{i=1}^{\infty} 2(i-1) q^{i-1} p + \sum_{i=1}^{\infty} q^{i-1} p$$

$$let\ j = i - 1$$

$$= \sum_{j=0}^{\infty} j^2 q^j p + 2 \sum_{j=0}^{\infty} j q^j p + 1 = qE[X^2] + 2qE[X] + 1$$

Other Discrete Probability Distributions
* The Geometric Random Variable

$$E[X^2] = qE[X^2] + 2qE[X] + 1, since\ E[X] = \frac{1}{p}$$

$$(1-q)E[X^2] = \frac{2q}{p} + 1 \Rightarrow pE[X^2] = \frac{2q+p}{p}$$

$$since\ p + q = 1 \Rightarrow E[X^2] = \frac{2q+p}{p^2} = \frac{q+1}{p^2}$$

$$Var[X] = E[X^2] - \left[E[X]\right]^2 = \frac{q+1}{p^2} - \frac{1}{p^2} = \frac{q}{p^2} = \frac{1-p}{p^2}$$

Other Discrete Probability Distributions
* The Geometric Random Variable

- **範例六**

假設今天要對你這個月的發票，已知在所有的發票中有N張的沒中獎發票與M張有中獎的發票。我們隨機拿出一張發票來對獎，試問:

(a) 需要n次才會拿到有中獎發票的機率為何?

(b) 至少需要k次才會拿到有中獎發票的機率為何?

Other Discrete Probability Distributions
* The Geometric Random Variable

Solution:

令X為抽到有中獎發票之前需要抽的次數。

抽到有中獎發票的機率為$p = \frac{M}{M+N}$

(a)

$$P\{X = n\} = \left(\frac{N}{M+N}\right)^{n-1} \frac{M}{M+N} = \frac{MN^{n-1}}{(M+N)^n}$$

(b)

$$P\{X \geq k\} = \frac{M}{M+N} \sum_{n=k}^{\infty} \left(\frac{N}{M+N}\right)^{n-1} = \frac{\frac{M}{M+N}\left(\frac{N}{M+N}\right)^{k-1}}{\left[1 - \frac{N}{M+N}\right]} = \left(\frac{N}{M+N}\right)^{k-1}$$

Other Discrete Probability Distributions
* The Negative Binomial Random Variable

- **The Negative Binomial Random Variable (負二項式分配)**
- 假設進行數個獨立試驗，成功的機率為$0 < p < 1$，試驗要一直持續到r次成功為止，令X為需要進行試驗的次數:

$$f(n; r, p) = P\{X = n\} = \binom{n-1}{r-1} p^r (1-p)^{n-r}, n = r, r+1, \ldots$$

- 換個角度思考，因為第n次一定要完成r次成功；在這之前，一定需要進行$n-1$次試驗，其中會有$r-1$次的成功。對於第一個事件的機率為

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

- 第二次的機率為p。

Other Discrete Probability Distributions
* The Negative Binomial Random Variable

- **Proof of expected value of negative binomial random variable**

$$E[X^k] = \sum_{n=r}^{\infty} n^k \binom{n-1}{r-1} p^r (1-p)^{n-r}$$

$$= \frac{r}{p} \sum_{n=r}^{\infty} n^{k-1} \binom{n}{r} p^{r+1} (1-p)^{n-r}, since \binom{n-1}{r-1} = \frac{r}{n} \binom{n}{r}$$

$let\ m = n+1$

$$= \frac{r}{p} \sum_{m=r+1}^{\infty} (m-1)^{k-1} \binom{m-1}{r} p^{r+1} (1-p)^{m-(r+1)} = \frac{r}{p} E[(Y-1)^{k-1}]$$

$where\ Y\ is\ negative\ binomial\ random\ variable\ with\ parameters\ r+1$

When $k = 1,\ E[X] = \frac{r}{p}$

Other Discrete Probability Distributions
* The Negative Binomial Random Variable

- **Proof of variance of negative binomial random variable**

$$E[X^k] = \frac{r}{p}E[(Y-1)^{k-1}]$$

When $k = 2$,

$$\frac{r}{p}E[(Y-1)^{2-1}] = \frac{r}{p}E[Y-1] = \frac{r}{p}\left(\frac{r+1}{p} - 1\right)$$

$$Var[X] = \frac{r}{p}\left(\frac{r+1}{p} - 1\right) - \left(\frac{r}{p}\right)^2 = \frac{r(1-p)}{p^2}$$

Other Discrete Probability Distributions
* The Negative Binomial Random Variable

- 範例七

如果進行一連串的獨立試驗，成功的機率為p，試問需要在失敗m次之前，成功r次之前的機率為何?

Solution:

在m次失敗之前需要達成r次成功，所以總試驗數為$r+m-1$次，故:

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

Other Discrete Probability Distributions
* The Negative Binomial Random Variable

- **範例八**

試求出投擲四次骰子點數為3，所需的次數期望值與變異數。

Solution:

因為要擲出四次點數為3，故為負二項式分配$(r = 4\ and\ p = \frac{1}{6})$。

$$E[X] = \frac{r}{p} = 24$$

$$Var[X] = \frac{r(1-p)}{p^2} = \frac{4(1-\frac{1}{6})}{\left(\frac{1}{6}\right)^2} = \frac{\frac{20}{6}}{\left(\frac{1}{6}\right)^2} = 120$$

Other Discrete Probability Distributions
* The Hypergeometric Random Variable

- **The Hypergeometric Random Variable (超幾何分配)**

假設總共有N個球在甕中，我們取出n顆不放回，其中有m顆是白色與$N-m$顆是黑色的。令X為白球被抽中的數量:

$$f(i;n,m,N)=P\{X=i\}=\frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}, where\ i=0,1,2,\ldots,n$$

如果覺得有點抽象，你可以想成有N個樣本，其中K個是不及格的。超幾何分布描述了在該N個樣本中抽出n個，其中k個是不及格的機率。

$$f(k;n,K,N)=P\{X=k\}=\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

Other Discrete Probability Distributions
* The Hypergeometric Random Variable

- **Proof of expected value of hypergeometric random variable**

$$E[X^k] = \sum_{i=0}^{n} i^k P\{X = i\} = \sum_{i=0}^{n} i^k \frac{\binom{m}{i}\binom{N-m}{m-i}}{\binom{N}{n}}$$

$$since\ i\binom{m}{i} = m\binom{m-1}{i-1}\ and\ n\binom{N}{n} = N\binom{N-1}{n-1}$$

$$E[X^k] = \frac{nm}{N}\sum_{i=1}^{n} i^{k-1} \frac{\binom{m-1}{i-1}\binom{N-1}{n-i}}{\binom{N}{n}}$$

$$= \frac{nm}{N}\sum_{j=0}^{n-1} (j+1)^{k-1} \frac{\binom{m-1}{j}\binom{N-1}{n-1-j}}{\binom{N-1}{n-1}} = \frac{nm}{N} E[(Y+1)^{k-1}]$$

$\boldsymbol{Y}$ is hypergeometric random variable with $n-1, N-1, m-1$.

Expected value

$When\ k = 1,$

$E[X] = nm/N$

Other Discrete Probability Distributions
* The Hypergeometric Random Variable

- **Proof of variance of hypergeometric random variable**

$$E[X^k] = \frac{nm}{N} E[(Y+1)^{k-1}]$$

$$E[X^2] = \frac{nm}{N} E[Y+1] = \frac{nm}{N}\left[\frac{(n-1)(m-1)}{N-1} + 1\right]$$

$$Var[X] = \frac{nm}{N}\left[\frac{(n-1)(m-1)}{N-1} + 1\right] - \left(\frac{nm}{N}\right)^2$$

$$= \frac{nm}{N}\left[\frac{(n-1)(m-1)}{N-1} + 1 - \frac{nm}{N}\right]$$

Other Discrete Probability Distributions
* The Hypergeometric Random Variable

$$Var[X] = \frac{nm}{N}\left[\frac{(n-1)(m-1)}{N-1} + 1 - \frac{nm}{N}\right]$$

$$since\ p = \frac{m}{N},$$

$$\frac{m-1}{N-1} = \frac{Np-1}{N-1} = p - \frac{1-p}{N-1}$$

$$Var[X] = np\left[(n-1)\left[p - \frac{1-p}{N-1}\right] + 1 - np\right]$$

$$= np\left[(n-1)p - (n-1)\frac{1-p}{N-1} + 1 - np\right] = np(1-p)(1 - \frac{n-1}{N-1})$$

If N is large ➔ $Var[X] \approx np(1-p) \sim Binomial$

Other Discrete Probability Distributions
* The Hypergeometric Random Variable

$$f(i;n,m,N)=P\{X=i\}=\frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}$$

- **範例九**

假設你是五金工廠的品管，你負責的產品是一包十個一組的螺絲。如果隨機抽檢3個都是正常的，就視這包螺絲沒有問題。假設一包螺絲中有4個壞掉的機率為30%，只有一個壞掉的機率為70%，試問因超過三個壞掉列為不良品螺絲包的比例會是多少?

Other Discrete Probability Distributions
* The Hypergeometric Random Variable

Solution:

$$f(i;n,m,N) = P\{X=i\} = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}$$

令A為該包螺絲為良品的事件:

$P(A)$

$$= P(A|lot\ has\ 4\ defectives) \times \left(\frac{3}{10}\right) + P(A|lot\ has\ 1\ dfective) \times \left(\frac{7}{10}\right)$$

$$= \frac{\binom{4}{0}\binom{6}{3}}{\binom{10}{3}} \times \left(\frac{3}{10}\right) + \frac{\binom{1}{0}\binom{9}{3}}{\binom{10}{3}} \times \left(\frac{7}{10}\right), since\ f(i=0;n=3,m,N=10)$$

$$= \frac{54}{100} \Rightarrow Hence, rejection\ rate = 1 - 0.54 = 46\%$$

Expected Value of Sums of Random Variables

- 期望值的一個非常重要的性質是隨機變量之和的期望值等於它們的期望之和。我們將在樣本空間 S為有限的或可數無限的假設下證明這一結果。 儘管沒有這個假設結果是正確的，但這個假設不僅會簡化論證，而且還會產生一個啟發性的證明。
- 假設我們有一個隨機變數X，令$X(s)$是當$s \in S$時實驗結果為X。如果X與Y都是隨機變數，則$Z = X + Y$中的Z也是隨機變數，也可以再表示為$Z(s) = X(s) + Y(s)$。

Expected Value of Sums of Random Variables

- **範例十**

如果今天你投擲一枚公平的硬幣五次，令X為前三次出現為正面的次數，Y為最後兩次出現正面的次數。

使得$Z = X + Y$，假設結果為$s = \{h, t, h, t, h\}$,

$$X(s) = 2$$

$$Y(s) = 1$$

$$Z(s) = X(s) + Y(s) = 2 + 1 = 3$$

Expected Value of Sums of Random Variables

令$p(s) = P(\{s\})$為實驗結果s所發生的機率，令事件A為有限或可數無限的互斥事件(mutually exclusive events)。則根據機率公理(Axioms of Probability)可以得知:

$$P(A) = \sum_{s \in A} p(s), when\ A = S, then \Rightarrow 1 = \sum_{s \in S} p(s)$$

令X為一隨機變數，期望值為$E[X]$。$X(s)$的數值為實驗結果為s的時候，因此$E[X]$可以被視為所有可能的X加權平均。

Expected Value of Sums of Random Variables

Proposition 1

$$E[X] = \sum_{s \in S} X(s)p(s)$$

Proof:

Suppose that the distinct values of X are $x_i, i \geq 1$. For each i, let S_i be the event that X is equal to x_i. That is, $S_i = \{s: X(s) = x_i\}$. Then,

$$E[X] = \sum_i x_i P\{X = x_i\} = \sum_i x_i \, P(S_i) = \sum_i x_i \sum_{s \in S_i} p(s) = \sum_i \sum_{s \in S_i} x_i p(s)$$

$$= \sum_i \sum_{s \in S_i} X(s)p(s)$$

Expected Value of Sums of Random Variables

$$= \sum_{i} \sum_{s \in S_i} X(s) p(s)$$

$since\ S_1, S_2, \ldots, are\ mutually\ exclusive\ events\ whose\ union\ is\ S.$

$$= \sum_{s \in S} X(s) p(s)$$

Expected Value of Sums of Random Variables

- **範例十一**

假設我們今天投擲硬幣兩次且為互相獨立，出現正面的機率為p，令X為出現正面的次數:

$$P(X=0)=P(t,t)=(1-p)^2$$
$$P(X=1)=P(h,t)+P(t,h)=2p(1-p)$$
$$P(X=2)=P(h,h)=p^2$$
$$E[X]=0\times(1-p)^2+1\times 2p(1-p)+2\times p^2=2p$$

Which agrees with

$$E[X]=X(h,h)p^2+X(h,t)p(1-p)+X(t,h)(1-p)p+X(t,t)(1-p)^2$$
$$=2p^2+p(1-p)+(1-p)p$$
$$=2p$$

Expected Value of Sums of Random Variables

- **Corollary 1**

For random variables $X_1, X_2, \dots, X_n$

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$$

Proof:

Let $Z = \sum_{i=1}^{n} X_i$. Then, by **Proposition 1** (pp.38).

$$E[Z] = \sum_{s \in S} Z(s)p(s) = \sum_{s \in S} (X_1(s) + X_2(s) + \cdots + X_n(s))p(s)$$

$$= \sum_{s \in S} X_1(s)p(s) + \sum_{s \in S} X_2(s)p(s) + \cdots + \sum_{s \in S} X_n(s)p(s)$$

$$= E[X_1] + E[X_2] + \cdots + E[X_n]$$

Expected Value of Sums of Random Variables

- **範例十二**

試問投擲n次骰子的點數和的期望值。

Solution:

令X為總和，故我們要求的是$E[X]$，X被定義為

$$X = \sum_{i=1}^{n} X_i$$

$$E[X_i] = \sum_{i=1}^{6} i\left(\frac{1}{6}\right) = \frac{21}{6} = \frac{7}{2}, therefore, E[X] = E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i] = 3.5n$$

Expected Value of Sums of Random Variables

- 範例十三

n次試驗有i次是成功，其成功的機率為$p_i, where\ i = 1, \dots, n$。
試問總成功次數期望值為何?

Solution:

Letting

$$X_i = \begin{cases} 1, if\ trial\ i\ is\ a\ success \\ 0, if\ trial\ i\ is\ a\ failure \end{cases}$$

$$X = \sum_{i=1}^{n} X_i \Rightarrow E[X] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} p_i$$

Expected Value of Sums of Random Variables

$$E[X] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} p_i$$

Note that this result does not require that the trials be independent.

Special Case 1: Binomial random variable

Assumes independent trials and all $p_i = p$, thus has mean np.

Special Case 2: Hypergeometric random variable

[加分題] What is the expected value?

Properties of the Cumulative Distribution Function

For the distribution function F of X, $F(b)$ denotes the probability that the random variable X takes on a value that is less than or equal to b. Following are some properties of the cumulative distribution function (CDF) F:

1. F is a nondecreasing function; that is, if $a < b$, then $F(a) \leq F(b)$
2. $\lim_{b \to \infty} F(b) = 1$
3. $\lim_{b \to -\infty} F(b) = 0$
4. F is right continuous. That is, for any b and any decreasing sequence $b_n, n \geq 1$, that converges to b, $\lim_{n \to \infty} F(b_n) = F(b)$.

Properties of the Cumulative Distribution Function

All probability questions about X can be answered in terms of the CDF (c.d.f.), F.

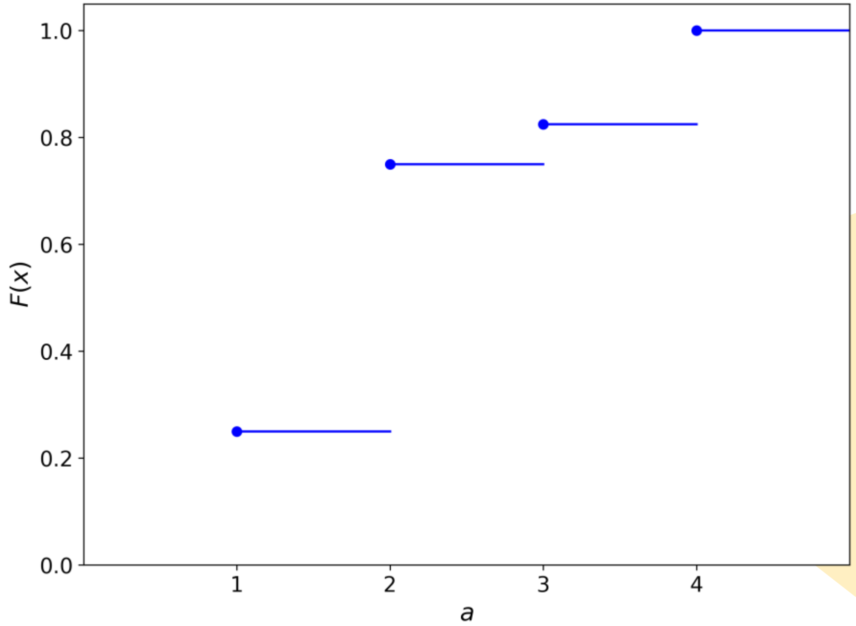
$$P\{a < X \leq b\} = F(b) - F(a) \; for \; all \; a < b$$

$$\{X \leq b\} = \{X \leq a\} \cup \{a < X \leq b\}$$

$$\{X \leq b\} = \{X \leq a\} + \{a < X \leq b\}$$

$$P\{X < b\} = P\left(\lim_{n\to\infty} \left\{X \leq b - \frac{1}{n}\right\}\right)$$

$$= \lim_{n\to\infty} P(X \leq b - \frac{1}{n}) = \lim_{n\to\infty} F(b - \frac{1}{n})$$



Properties of the Cumulative Distribution Function

- **範例十四**

隨機變數X的累積機率質量函數為

$$F(x) = \begin{cases} 0, x < 0 \\ \frac{x}{2}, 0 \leq x < 1 \\ \frac{2}{3}, 1 \leq x < 2 \\ \frac{11}{12}, 2 \leq x < 3 \\ 1, 3 \leq x \end{cases}$$

試問: (a) $P\{X < 3\}$; (b) $P\{X = 1\}$; (c) $P\{X > \frac{1}{2}\}$; (d) $P\{2 < X \leq 4\}$

Properties of the Cumulative Distribution Function

Solution:

(a) $P\{X < 3\}$; **(b)** $P\{X = 1\}$; **(c)** $P\{X > \frac{1}{2}\}$; **(d)** $P\{2 < X \leq 4\}$

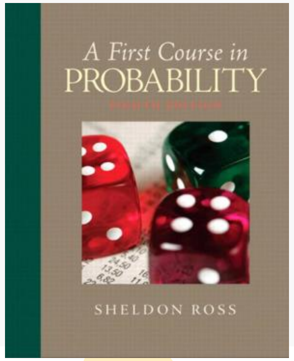
(a) $P\{X < 3\} = \lim_{n}\{X \leq 3 - \frac{1}{n}\} = \lim_{n} F\left(3 - \frac{1}{n}\right) = \frac{11}{12}$

(b) $P\{X = 1\} = P\{X \leq 1\} - P(X < 1) = F(1) - \lim_{n} F\left(1 - \frac{1}{n}\right) = \frac{2}{3} - \frac{1}{2} = \frac{1}{6}$

(c) $P\left\{X > \frac{1}{2}\right\} = 1 - P\left\{X \leq \frac{1}{2}\right\} = 1 - F\left(\frac{1}{2}\right) = \frac{3}{4}$

(d) $P\{2 < X \leq 4\} = F(4) - F(2) = \frac{1}{12}$

[#9] Assignment



- Selected Problems from Sheldon Ross Textbook [1].

4.21. Four buses carrying 148 students from the same school arrive at a football stadium. The buses carry, respectively, 40, 33, 25, and 50 students. One of the students is randomly selected. Let X denote the number of students that were on the bus carrying the randomly selected student. One of the 4 bus drivers is also randomly selected. Let Y denote the number of students on her bus.

(a) Which of $E[X]$ or $E[Y]$ do you think is larger? Why?

(b) Compute $E[X]$ and $E[Y]$.

4.38. If $E[X] = 1$ and $\mathrm{Var}(X) = 5$, find

(a) $E[(2 + X)^2]$;

(b) $\mathrm{Var}(4 + 3X)$.

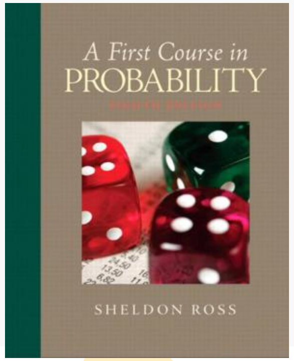
4.40. On a multiple-choice exam with 3 possible answers for each of the 5 questions, what is the probability that a student will get 4 or more correct answers just by guessing?

4.48. It is known that diskettes produced by a certain company will be defective with probability .01, independently of each other. The company sells the diskettes in packages of size 10 and offers a money-back guarantee that at most 1 of the 10 diskettes in the package will be defective. The guarantee is that the customer can return the entire package of diskettes if he or she finds more

4.51. The expected number of typographical errors on a page of a certain magazine is .2. What is the probability that the next page you read contains (a) 0 and (b) 2 or more typographical errors? Explain your reasoning!

[1] Sheldon Ross. A First of Course in Probability. 8th edition.

[#9] Assignment



- Selected Problems from Sheldon Ross Textbook [1].

4.52. The monthly worldwide average number of airplane crashes of commercial airlines is 3.5. What is the probability that there will be

(a) at least 2 such accidents in the next month;

(b) at most 1 accident in the next month?

Explain your reasoning!

4.57. Suppose that the number of accidents occurring on a highway each day is a Poisson random variable with parameter $\lambda = 3$.

(a) Find the probability that 3 or more accidents occur today.

(b) Repeat part (a) under the assumption that at least 1 accident occurs today.

4.64. The suicide rate in a certain state is 1 suicide per 100,000 inhabitants per month.

(a) Find the probability that, in a city of 400,000 inhabitants within this state, there will be 8 or more suicides in a given month.

(b) What is the probability that there will be at least 2 months during the year that will have 8 or more suicides?

(c) Counting the present month as month number 1, what is the probability that the first month to have 8 or more suicides will be month number $i, i \geq 1$?

What assumptions are you making?

[1] Sheldon Ross. A First of Course in Probability. 8th edition.

Reference

Ross, S. (2010). *A first course in probability*. Pearson.

The End

If you have any questions, please do not hesitate to ask me.

Thank you for your attention))