CSE 107 2-3-25

- Mid 1 Solutions Posted
- lab3 Posted
- hw 4 Posted

Independence of 3 or more r.v.s

We say X, Y, Z are indep. iff

$$p_{X,Y,Z}(x,y,z) = p_X(x) \cdot p_Y(y) \cdot p_Z(z)$$

for all $x \in range(X)$, $y \in range(Y)$ and $z \in range(Z)$.

Also, if X, Y, Z are indep., then so are $g(X), f(Y), h(Z)$.

Also, X, Y, Z indep. implies

$$\boxed{Var(X+Y+Z) = Var(X) + Var(Y) + Var(Z)}$$

Exercise Prove this

Variance of the Binomial r.v.

Recall that a Binomial r.v. with Param. n, p is a sum

$$X = X_1 + X_2 + \cdots + X_n$$

of n indep. Bernoulli r.v.s with param. p.

Thus

by independence

$$\mathrm{Var}(X) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n)$$

$$= p(1-p) + \cdots + p(1-p)$$

$$= np(1-p).$$

Suppose we have a weighted coin with

$$P(\text{Head}) = p$$

is unknown. How do we estimate p?

We compute the <u>Sample Mean</u>

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

where $X_i = \begin{cases} 1 & \text{if head (on } i^{th} \text{ flip)} \\ 0 & \text{if tail} \end{cases}$. We have

$$E[X_i] = p$$

so

$$E[S_n] = E\left[\tfrac{1}{n}X_1 + \tfrac{1}{n}X_2 + \cdots + \tfrac{1}{n}X_n\right]$$

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

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

Thus S_n serves as an approx. to p.
How good is the estimate?
We have

$$Var(S_n) = Var\Big(\sum_{i=1}^{n} \underbrace{\tfrac{1}{n} X_i}_{\text{indep.}}\Big)$$

$$= \sum_{i=1}^{n} Var\left(\frac{1}{n} \cdot X_i\right)$$

$$= \cancel{n} \cdot \frac{1}{n^{\cancel{2}}}\, p(1-p)$$

$$= \frac{p(1-p)}{n} \longrightarrow 0 \quad \text{as } n \to \infty$$

so $Var(\hat{p}_n) \longrightarrow 0$

<u>Exercise</u>

Prove that if $Var(Y) = 0$, then

$Y = \text{const.} = E[Y]$

Note

even if X_i are not Bernolli, but still independent and have same pmf as X, then

$$Var(S_n) = \frac{Var(X)}{n}$$

so $Var(S_n) \rightarrow 0$, hence

$$E[S_n] = E[X]$$

so S_n is a good approx. to $E[X]$.

Ex. (# 38 on p. 132)

Alice passes through 4 traffic lights on way to work, each equally likely to be red or green, independently.

a) find PMF, mean, Var of the # of red lights.

Soln

Let X = # red lights. Then X is binomial parameters $n = 4$, and $p = \frac{1}{2}$.

Thus

$$p_X(k) = \binom{4}{k} \cdot \left(\tfrac{1}{2}\right)^k \left(\tfrac{1}{2}\right)^{4-k} \qquad (k = 0,1,2,3,4)$$

$$= \binom{4}{k}\left(\tfrac{1}{2}\right)^4$$

$$= \frac{4!}{k!\,(4-k)!} \cdot \frac{1}{16} = \frac{24}{16 \cdot k!\,(4-k)!}$$

$$= \boxed{\frac{3}{2 \cdot k!\,(4-k)!} \qquad (k = 0, \ldots, 4)}$$

$$E[X] = np = 4 \cdot \tfrac{1}{2} = \boxed{2}$$

$$Var(X) = np(1-p) = 4 \cdot \tfrac{1}{2} \cdot \tfrac{1}{2} = \boxed{1}$$

b.] Suppose Alice is delayed by 2 minutes by each red light. Find the variance of her commute time.

Soln

Let T be commute time, and T_0 her commute time with no red lights. Then

$$T = T_0 + 2X \quad (\text{minutes})$$

so

$$Var(T) = Var(2X) = 2^2 \cdot Var(X) = \boxed{4}$$

3.1 Continuous Random Variables

Defn

A random variable $X : \Omega \to \mathbb{R}$ is called continuous iff there exists a function

$$f_X : \mathbb{R} \to \mathbb{R}$$

called the probability density function (PDF) such that

$$P(X \in I) = \int_I f_X(x)\,dx$$

for 'any' subset $I \subseteq \mathbb{R}$.

In particular, if $I = [a,b]$ is an interval, then

$$P(a \le X \le b) = \int_a^b f_X(x)\,dx .$$

<u>note</u>

integrating over $[a,b]$, $[a,b)$, $(a,b]$ or (a,b) are the same, since

$$P(X=a) = \int_a^a f_X(x)\,dx = 0$$

Note:

to be a valid PDF, must have

$$\int_{-\infty}^{\infty} f_X(x)\,dx = P(X \in \mathbb{R}) = P(\Omega) = 1$$