CSE 107 2-2-24

Supplemental Lecture

Theorem

If random variables X, Y are independent, then

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

Proof.

Let X, Y be indep. r.v.s. Set

$$U = X - E[X]$$

and

$$V = Y - E[Y]$$

Then $E[U] = E[X] - E[X] = 0$, and likewise $E[V] = 0$. Also

$$
\begin{aligned}
U + V &= X - E[X] + Y - E[Y] \\
&= (X+Y) - (E[X] + E[Y]) \\
&= (X+Y) - E[X+Y]
\end{aligned}
$$

Thus

$$
\begin{aligned}
Var(X+Y) &= E[(U+V)^2] \\
&= E[U^2 + 2UV + V^2] \\
&= E[U^2] + 2E[UV] + E[V^2] \\
&= E[U^2] + 2\cancel{E[U]}^{0} \cdot \cancel{E[V]}^{0} + E[V^2] \\
&= Var(X) + Var(Y).
\end{aligned}
$$

■

what we've done for 2 r.v.s, we can do for 3 or more

- let X, Y, Z be r.v.s, we say they are independent iff

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

for all x, y, z

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

- if X, Y, Z indep., then

$$E[XYZ] = E[X] \cdot E[Y] \cdot E[Z]$$

and

$$Var(X+Y+Z) = Var(X) + Var(Y) + Var(Z).$$

Variance of the Binomial: n, p

Recall: X (Binomial: n, p) is a sum

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

where each X_i is Bernoulli with param. p.

<u>note</u>: $\{X_1, X_2, \ldots, X_n\}$ are independent.

Thus

$$\mathrm{Var}(X) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_n)$$

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

$$= n p(1-p)$$

∎

Ex

Suppose we have a weighted coin, but we don't know $p = P(\text{Head})$. how can we estimate p ?

flip n times, see what happens.

compute sample mean S_n

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

where $X_i = 1$ or 0 depending on i^{th} flip (H or T). Also called relative frequency of Heads.

since p is the true Prob. of heads,

$$E[X_i] = p \quad (1 \le i \le n)$$

so

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

$$= \frac{1}{n}\underbrace{E[X_1]}_{p} + \frac{1}{n}\underbrace{E[X_2]}_{p} + \cdots + \frac{1}{n}\underbrace{E[X_n]}_{p}$$

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

Thus S_n serves as an estimate of $E[S_n] = p$. Since the X_i are indep., so are $\frac{1}{n} \cdot X_i$, hence

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

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

$$= \sum_{i=1}^{n} \frac{1}{n^2} \underbrace{Var(X_i)}_{p(1-p)}$$

$$= \sum_{i=1}^{n} \frac{1}{n^2} \cdot p(1-p)$$

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

$$= \frac{p(1-p)}{n}$$

observe: $\lim_{n\to\infty} Var(S_n) = \lim_{n\to\infty} \frac{p(1-p)}{n} = 0$

so S_n becomes a good approximation to $E[S_n] = p$ for large n.

<u>note</u>: even if X_i are not Bernoulli, the same calculation gives

$$Var(S_n) = \frac{Var(X_i)}{n} \longrightarrow 0$$

provided X_i are independent.

so since $E[S_n] = E[X_i]$, we have S_n is a good approx. to $E[X_i]$

<u>Ex.</u> (#38 on p. 132 ch. 2)

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

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

let $X = \#$ red lights. X is Binomial with param: $n=4$, $p=\frac{1}{2}$.

so

$$p_X(k) = \binom{4}{k} \cdot \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{4-k}$$

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

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

$$= \begin{cases} \frac{3}{2} \cdot \frac{1}{k!(4-k)!} & \text{if } k = 0, 1, 2, 3, 4 \\ 0 & \text{otherwise} \end{cases}$$

$$E[X] = n \cdot p = 4 \cdot \frac{1}{2} = \boxed{2}$$

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

(b) Suppose Alice is delayed by 2 minutes by each red light. Find Variance of her commute time.

Let $T = T_0 + 2X$ where T_0 is her commute time with no red lights. So T is her commute time.

Then

$$\begin{aligned} Var(T) &= Var(T_0 + 2X) \\ &= Var(2X) \\ &= 4\,Var(X) \\ &= 4 \cdot 1 = \boxed{4} \end{aligned}$$