Class 10: Variance & Joint PMFs

Thursday, October 3, 2019 11:39 AM

- Midterm 1 on Tues, Oct 8, 11:40-12:55, in Schermerhorn 614 and CEPSR 750.
  (Room assignments to come.)

Recap:

Sample space $= \Omega$

$X : \Omega \to \mathbb{R}$

$p : \mathbb{R} \to [0,1]$ $\quad p(k) = \mathbb{P}(X=k)$

$\sum_k p(k) = 1$.

Let X be a discrete r.v.

Let p is its PMF.

Defined $\mathbb{E}[X] = \sum_k k\, p(k)$.

More generally $\mathbb{E}[f(X)] = \sum_k f(k)\, p(k)$.

Ex: $\mathbb{E}[X^2] = \sum_k k^2 p(k)$.

Defined $Var(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$ measures "spread".

Ex: If $\mathbb{E}[X] = 2$, then $Var(X) = \mathbb{E}[(X-2)^2]$

$= \sum_k (k-2)^2 p(k)$.

Also $Sd(X) = \sqrt{Var(X)}$.

Properties: Let $a, b \in \mathbb{R}$, constants.

(1) Linearity: $\mathbb{E}[aX+b] = a\mathbb{E}[X] + b$.

(2) $Var(aX+b) = a^2 Var(X) \longrightarrow Sd(aX+b) = |a|\, Sd(X)$

shifting does not effect variance!

Proof: Let $m = \mathbb{E}[X]$. Then $\mathbb{E}[aX+b] \overset{(1)}{=} am+b$.

Then

$$Var(aX+b) = \mathbb{E}\left[\left((aX+b) - (am+b)\right)^2\right]$$

$$= \mathbb{E}\left[(aX - am)^2\right]$$

$$= \mathbb{E}[(aX - am)^2]$$
$$= \mathbb{E}[a^2(X-m)^2]$$
$$\overset{①}{=} a^2\,\mathbb{E}[(X-m)^2]$$
$$= a^2\,\mathrm{Var}(X).$$

Ex: $a=-1$, $b=0$ shows $\mathrm{Var}(X) = \mathrm{Var}(-X)$.

③ $\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$

Proof: Let $m = \mathbb{E}[X]$. Definition of variance says

$$\mathrm{Var}(X) = \mathbb{E}[(X-m)^2]$$
$$= \mathbb{E}[X^2 - 2mX + m^2]$$
$$\overset{①}{=} \mathbb{E}[X^2] - 2m\,\underbrace{\mathbb{E}[X]}_{m} + m^2 \quad \text{since } m \text{ is constant}$$
$$= \mathbb{E}[X^2] - 2m^2 + m^2$$
$$= \mathbb{E}[X^2] - m^2$$
$$= \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

"with probability"

Ex: $X = \begin{cases} 1 & \text{w.p. } \frac{1}{2} \\ -1 & \text{w.p. } \frac{1}{2} \end{cases}$

$\mathbb{E}[X] = 0$

$\mathrm{Var}(X) = \mathbb{E}[(X-0)^2]$
$= \mathbb{E}[X^2] = 1.$

So $\mathbb{E}[X^2] \neq (\mathbb{E}[X])^2$.

Note: $\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] \geq 0$, so

③ implies $\mathbb{E}[X^2] \geq (\mathbb{E}[X])^2$.

This is strict inequality always, unless X is non-random.

Poisson: Let $X \sim \mathrm{Poiss}(\lambda)$, $\lambda > 0$. This means X has PMF $p(k) = e^{-\lambda} \frac{\lambda^k}{k!}$ for $k = 0, 1, 2, \dots$.

Goal 1: Find $\mathbb{E}[X]$.

$$\mathbb{E}[X] = \sum_k k\, p(k) = \sum_{k=0}^{\infty} k\, e^{-\lambda} \frac{\lambda^k}{k!}$$

$$= e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!}$$ since $k=0$ term gives 0

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

$$= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}$$ $j = k-1$

$$= \lambda e^{-\lambda} \underbrace{\sum_{j=0}^{\infty} \frac{\lambda^j}{j!}}_{= e^{\lambda}}$$ Taylor series

Alternatively:

$$= \lambda \sum_{j=0}^{\infty} e^{-\lambda} \frac{\lambda^j}{j!} = \lambda \underbrace{\sum_{j=0}^{\infty} p(j)}_{=1 \text{ since probabilities add up to } 1} = \lambda$$

$$= \lambda e^{-\lambda} \cdot e^{\lambda}$$

$$= \lambda.$$

Goal 2: Find $Var(X)$.

Use $Var(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \mathbb{E}[X^2] - \lambda^2$.

So let's find $\mathbb{E}[X^2]$.

Write $\mathbb{E}[X^2] = \mathbb{E}[X(X-1) + X]$

$$= \mathbb{E}[X(X-1)] + \mathbb{E}[X]$$ linearity of $\mathbb{E}$

$$= \mathbb{E}[X(X-1)] + \lambda.$$

So let's find $\mathbb{E}[X(X-1)]$.

$$\mathbb{E}[X(X-1)] = \sum_k k(k-1)\, p(k) = \sum_{k=0}^{\infty} k(k-1)\, e^{-\lambda} \frac{\lambda^k}{k!}$$

$$= e^{-\lambda} \sum_{k=2}^{\infty} k(k-1) \frac{\lambda^k}{k!}$$ since $k=0$ and $k=1$ give 0

λ^k

$$= e^{-\lambda} \sum_{k=2} k(k-1) \frac{\lambda^k}{k!}$$ since $k=0$ and $k=1$ give 0

$$= e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^k}{(k-2)!}$$

$$= \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!}$$

$j = k-2$

$$= \lambda^2 \underbrace{e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!}}_{\textstyle \sum_{j=0}^{\infty} p(j) = 1}$$

$$= \lambda^2.$$

Putting it together: $\mathbb{E}[X(X-1)] = \lambda^2$

$$\mathbb{E}[X^2] = \mathbb{E}[X(X-1)] + \lambda = \lambda^2 + \lambda$$

$\hookrightarrow$ $Var(X) = \mathbb{E}[X^2] - \lambda^2 = \lambda^2 + \lambda - \lambda^2$

$Var(X) = \lambda$.

Midterm Recap. (up to 10/1)

- Sample Spaces: events & outcomes
  - Sample Spaces $\rightarrow \Omega$
  - events $\rightarrow$ subset of Ω
  - outcomes $\rightarrow$ element of the set Ω
  - Union, intersection, complement
- Rules/axioms of probability & conditional probability:
  - $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \underbrace{\mathbb{P}(A \cap B)}_{0 \text{ if } A, B \text{ disjoint}}$
  - $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
  - $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ conditional prob. of A given B

- Bayes rule $\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}$

- Law of total probability
$$\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)$$

- Independence:
  - 2 events: $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) \longleftrightarrow \mathbb{P}(A|B) = \mathbb{P}(A)$
  - ≥ 3 events: joint independence
    ↳ probabilities of any group factorizes into the product of individual probabilities

- Counting Problems (equally likely outcomes)
  - Here $\mathbb{P}(E) = \frac{\text{\# ways } E \text{ can happen}}{\text{total \# outcomes}}$.
  - Ordered / Unordered?
  - With or without replacement?

- Random Variables:
  - Definition $X : \Omega \to \mathbb{R}$
  - PMF (probability mass function) $p(k) = \mathbb{P}(X=k)$
    ↳ calculating probabilities like $\mathbb{P}(X \leq 3) = \sum_{k \leq 3} p(k)$
    or $\mathbb{P}(2 \leq X \leq 7) = \sum_{2 \leq k \leq 7} p(k)$
  - Transformations $Y = f(X)$ → find PMF of Y
    e.x. $Y = 2(X-2)^2$

- Common RVs: What do they model? What parameter(s)? What is PMF?

Bernoulli $\text{Ber}(p)$ $0 \le p \le 1$

- binary outcomes
- yes/no?
- success/failure?

$$\mathbb{P}(X=k) = \begin{cases} p & \text{if } k=1 \\ 1-p & \text{if } k=0 \end{cases}$$

Binomial $\text{Bin}(n,p)$ $0 \le p \le 1$, $n=0,1,\dots$

successes out of n indep. trials with success probability p

$$\mathbb{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k=0,1,\dots,n$$

Geometric $\text{Geom}(p)$ $0 \le p \le 1$

failures until first success in repeated trials

- $$\mathbb{P}(X=k) = p(1-p)^{k-1}$$

for $k=1,2,\dots$

if you count the 1st heads

- If you don't count 1st heads, slightly different:

$$\mathbb{P}(X=k) = p(1-p)^k$$

for $k=0,1,2,\dots$

Poisson $\text{Poiss}(\lambda)$ $\lambda > 0$

rare events

$$\text{Bin}\left(n, \tfrac{\lambda}{n}\right) \xrightarrow{n \to \infty}$$

$$\mathbb{P}(X=k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k=0,1,2,\dots$$

- Expectations:

$$\mathbb{E}[X] = \sum_k k\, p(k)$$

$$\mathbb{E}[f(X)] = \sum_k f(k)\, p(k)$$

(No variance.)