Lecture 8 - 9/9

Saturday, September 10, 2022 12:14 PM

Review:

① Independent random variables:

we say discrete random variables X and Y are independent if

$$P(X=A, Y=B) = P(X=A)\,P(Y=B) \quad \text{for any } A \text{ and } B.$$

Equivalently, X and Y are independent if

$$P(X=i, Y=j) = p_X(i)\,p_Y(j) \quad \text{for any } i \text{ and } j.$$

② Distributions: we only care about statistical properties, so we only focus on distribution of random variables.

③ Important distributions.

3.1 Bernoulli distribution: we say a random variable X has $Ber(p)$ distribution if $P(X=1)=p$ and $P(X=0)=1-p$.

- $E[X] = p \quad Var(X) = p(1-p)$

3.2 Binomial distribution: we say a random variable X has $Bi(n,p)$ distribution if:

$$p_X(k) = P(X=k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{for all } k \in \{0,1,\dots,n\}$$

- $E[X] = np$, $Var(X) = np(1-p)$
- Mode of PMF $= \lfloor (n+1)p \rfloor$. Median: $|\hat{k}_X - np| \leq \max\{p, 1-p\}$

Today:

①. Important statistical quantities

②. Geometric distribution

③. Bernoulli process and negative Binomial distribution

④. Poisson distributions

---

①. Important statistical quantities:

① Important statistical quantities:

As we discussed, all we care is the statistical properties of an experiment. This was our main motivation for defining random variables. Some important statistical quantities are: (we focus on discrete random variables)

(i) Expected value: it is the average number we expect to see as realization.

$$E[X] = \sum_k k\, p_X(k)$$

(ii) Variance: it measures how spread-out is the random variable:

$$Var(X) = E\left[(X - E[X])^2\right] = \sum_k (k - E[X])^2\, p_X(k)$$

(iii) Mode: the mode of pmf is the most probable outcome:

$$k^* = \arg\max_k p_X(k), \text{ equivalently } p_X(k^*) \geq p_X(k) \text{ for all } k.$$

in case there are multiple candidates for k^*, all of them are considered as mode.

(iv) Median: median of a random variable is the smallest number $\hat{k}$ in its support for which:

$$P(X \leq \hat{k}) \geq 0.5$$

or equivalently

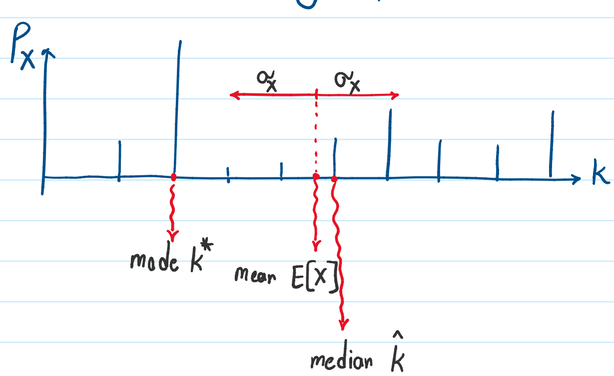
$$P(X \leq \hat{k}) \geq P(X > \hat{k})$$

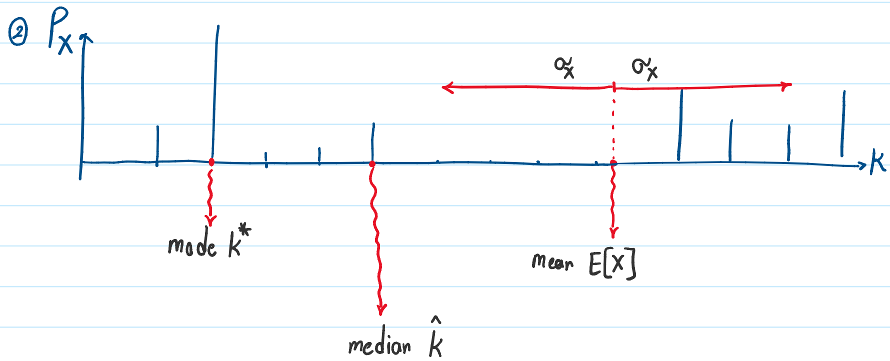
intuitively speaking, it is the number for which

$$P(X \leq \hat{k}) \approx P(X > \hat{k}) \approx 0.5$$

e.g. consider the following pmfs:

①



② 

Notice that

(*) median and mode only depend on pmf.

(**) Expected value and variance depend on pmf and values that X takes.

② Geometric random variables:

. Toss a coin till the first time H is observed.

$\Omega = \{\text{sequences of form } TT\cdots TH\}$, $X(\omega) =$ number of tosses

. Transmit a bit over a channel till a successful transmission.

$\Omega = \{\text{sequences of form } FFF\cdots FS\}$. $X(\omega) =$ number of transmissons.

In each of above cases, we are repeating independent $Ber(p)$ trials till we observe one:

$$P(k \text{ trialls till a success}) = \underbrace{(1-p)^{k-1}}_{\substack{\text{first } k-1 \\ \text{trials are } 0}} \underbrace{p}_{\substack{\text{last trial} \\ \text{is } 1}} \qquad k \in \{1,2,3,\dots\}$$

We say a random variable X has geometry distribution with paramcter p, (has $Geo(p)$ distribution) if

$$P_X(k) = (1-p)^{k-1} p \qquad k \in \{1,2,3,\dots\}$$

• We need to check P_X satisfies axioms of probability. In particular, we need to check $\sum_{k=1}^{\infty} P_X(k) = 1$. Recall that $(1-x)^{-1} = 1 + x + x^2 + \cdots$ for $|x| < 1$ which you can check by calculating $(1+x)(1-x)^{-1} = 1$. Using this, we have

$$\sum^{\infty} P_X(k) = \sum^{\infty} (1-p)^{k-1} p = p \sum^{\infty} (1-p)^{\ell} = p\,(1-(1-p))^{-1} = 1$$

which you can check by calculating $(1+x)(1-x) = 1$. Using this, we have

$$\sum_{k=1}^{\infty} P_X(k) = \sum_{k=1}^{\infty} (1-p)^{k-1} p = p \sum_{\ell=0}^{\infty} (1-p)^{\ell} = p\left(1-(1-p)\right)^{-1} = 1$$

- tail of Geo(p): $P(X > m) = P(\{\text{first } m \text{ trials is faliure}\}) = (1-p)^m$

- Mean of Geo(p) random variable: there are three ways to calculate $E[X]$.

Method 1. Notice that $(1-x)^{-1} = 1+x+x^2+\cdots$. Taking derivatives from both sides, we have

$$\frac{1}{(1-x)^2} = 1+2x+3x^2+\cdots$$

$$E[X] = \sum_{k=1}^{\infty} k\, P_X(k) = \sum_{k=1}^{\infty} k(1-p)^{k-1} p$$

$$= p \sum_{\ell=0}^{\infty} (\ell+1)(1-p)^{\ell} = p \frac{1}{(1-(1-p))^2} = \frac{1}{p}$$

Method 2: we can use a trick in calculation, writing $E[X]$ in terms of $E[X]$!

$$E[X] = \sum_{k=1}^{\infty} k\, P_X(k) = \sum_{k=1}^{\infty} k(1-p)^{k-1} p$$

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

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

$$= p + (1-p)\Big(\underbrace{\sum_{\ell=1}^{\infty} (1-p)^{\ell-1} p}_{=1} + \underbrace{\sum_{\ell=1}^{\infty} \ell(1-p)^{\ell-1} p}_{=E[X]} \Big)$$

$$= p + (1-p)(1 + E[X])$$

Method 3. Consider the continuation random variable after first trial! Define Y as

$$Y = \begin{cases} 0 & \rightsquigarrow \text{if first trial was success} \\ \text{number of trial to see sucess} & \rightsquigarrow \text{if first trial was faliure} \end{cases}$$

Notice that $X = Y+1$, and we have

$$P_Y(k) = \begin{cases} p & k=0 \\ (1-p)^k p & k=1,2,3,\dots \end{cases}$$

which implies:
$$E[Y] = \sum_{k=0}^{\infty} k\, P_Y(k) = \sum_{k=1}^{\infty} k\,(1-p)^k p = (1-p)\, E[X]$$

Using above equalities:
$$E[X] = 1 + E[Y] = 1 + (1-p)E[X] \Rightarrow E[X] = \frac{1}{p}$$

- Variance of Geo(p) random variable: we can replicate 3 methods above for variance as well.

We only consider the last method. Using $X = 1+Y$, we have
$$E[X^2] = E[(1+Y)^2] = E[1+Y^2+2Y] = 1+2E[Y]+E[Y^2]$$

We already know $E[Y] = (1-p)E[X] = \frac{1-p}{p}$. Hence, we need to calculate $E[Y^2]$ only:
$$E[Y^2] = \sum_{k=0}^{\infty} k^2 P_Y(k) = \sum_{k=1}^{\infty} k^2 (1-p)^k p = (1-p)\, E[X^2]$$

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

$$\rightsquigarrow \quad Var(X) = E[X^2] - (E[X])^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}$$

- Memoryless property of Geometric r.v.: Consider a Geo(p) random variable X. One of the important properties of X is

$$P(X > m+n \mid X > n) = \frac{P(\{X > m+n\} \cap \{X > n\})}{P(\{X > n\})}$$
$$= \frac{P(X > m+n)}{P(X > n)}$$
$$= \frac{(1-p)^{m+n}}{(1-p)^n} = (1-p)^m$$

$$\rightsquigarrow \quad P(X > m+n \mid X > n) = P(X > m)$$

$$\rightsquigarrow P(X>m+n \mid X>n) = P(X>m)$$

Consider a lamp, and suppose the lamp fails to work each day with probability p, independent of anything else. Notice that this assumption is unrealistic as we expect p to change over time. However, if p is fixed, the memoryless property says:

$$P(\text{lamp survives } m \text{ more days} \mid \text{lamp has already survived } n \text{ days})$$

$$= P(\text{lamp survives } m \text{ days}) \quad !!!$$

---

③ Bernoulli process:

A Bernoulli process is an infinite sequence of independent Bernoulli random variables with parameter p, i.e., infinite sequence of $Ber(p)$ independent random variables:

$$(X_1, X_2, X_3, \dots)$$

In particular,

① if we pick a time index i, X_i is a Bernoulli random variable.

② Notice that each X_i is a random variable, i.e., $X_i : \Omega \to \mathbb{R}$. So if we pick an $\omega \in \Omega$, the resulted sequence $(X_1(\omega), X_2(\omega), \dots)$ is a sequence of zeros and ones, which is called a sample path. In particular, a sample path is a function of time. For each $\omega \in \Omega$:

$$f_\omega : \{1,2,3,\dots\} \to \mathbb{R}, \quad f_\omega(K) = X_K(\omega)$$

e.g: consider a communicating channel, where the sender sends 0 and 1 at each time index, for eternity, and reciever recieves the same bit with probability p, and opposite bit with probability $1-p$.

Define:

$$X_i = \begin{cases} 1 & \text{successful transmission at time } i \\ 0 & \text{failed transmission at time } i \end{cases}$$

Consider a Bernoulli process:

- Let L_1 = number of trials till first one

Consider a Bernoulli process:

- Let L_1 = number of trials till first one

  L_2 = number of trials after L_1 till first one

  $\vdots$

  L_k = number of trials after $L_1 + \cdots + L_{k-1}$ till first one

  $\ddots$

L_i s are independent Geo(p) random variables. There is a one-to-one correspondence between $(X_1, X_2, X_3, \ldots)$ and $(L_1, L_2, L_3, \ldots)$

- Let C_n denote the number of ones in the first n trials. C_n is a Ber(n,p) random variable. In particular, $C_n = \sum_{i=1}^{n} X_i$, and C_i s are not independent from each other. However $C_n - C_m$ and C_m are independent for any $n > m$. There is a one-to-one correspondence between $(X_1, X_2, \ldots)$ and $(C_1, C_2, \ldots)$

- Let S_r denote number of trials till r'th one in the sequence. In particular,

  $S_1 = L_1$

  $S_2 = L_1 + L_2$

  $\vdots$

  $S_r = L_1 + \cdots + L_r$

  $\vdots$

S_r s are not independent from each other but $S_r - S_\ell$ is independent of S_ℓ for any $r > \ell$.

S_r has Negative Binomial distribution, which we will discuss next.

e.g: Consider a sample path $(X_1(\omega), X_2(\omega), X_3(\omega), \ldots) = (0,1,1,0,1,0,0,1,\ldots)$

- $(L_1(\omega), L_2(\omega), L_3(\omega), \ldots) = (2,1,2,3,\ldots)$
- $(C_1(\omega), C_2(\omega), C_3(\omega), \ldots) = (0,1,2,2,3,3,3,4,\ldots)$
- $(S_1(\omega), S_2(\omega), S_3(\omega), \ldots) = (2,3,5,8,\ldots)$

3.1 Negative binomial distribution.

. Toss a coin till r heads is observed. X = number of tosses.

$$P(\text{We need } n \text{ tosses}) = \binom{n-1}{r-1} \times p^r \times (1-p)^{n-r}$$

$$P(\text{We need } n \text{ tosses}) = \underbrace{\binom{n-1}{r-1}} \times \underbrace{p^r} \times \underbrace{(1-p)^{n-r}}$$

r ones happens in n trials
last trial is one, we need to
pick location of other trials.

give locations of ones
we include probability
that they are one

rest trials should be zero

Def. We say a random variable X has negative Binomial distribution, denoted by $NB(r,p)$ if:

$$P_X(n) = \binom{n-1}{r-1} p^r (1-p)^{n-r}, \quad n \geq r$$

- We need to show P_X satisfies the axioms. In particular, we need to show

$$\sum_{n=r}^{\infty} P_X(n) = 1$$

Notice that Maclaurin series (Taylor expansion around 0) of $(1-x)^{-r}$ is

$$(1-x)^{-r} = \sum_{k=0}^{\infty} \binom{k+r-1}{r-1} x^k$$

which can be derived using:

$$\frac{d}{dx}(1-x)^{-r} = r(1-x)^{-r-1} \longrightarrow \left.\frac{d}{dx}(1-x)^{-r}\right|_{x=0} = r$$

$$\frac{d^2}{dx}(1-x)^{-r} = r(r+1)(1-x)^{-r-2} \longrightarrow \left.\frac{d^2}{dx^2}(1-x)^{-r}\right|_{x=0} = r(r+1)$$

$$\vdots$$

$$\frac{d^k}{dx}(1-x)^{-r} = r(r-1)\cdots(r-k+1)(1-x)^{-r-k} \longrightarrow \left.\frac{d^k}{dx^k}(1-x)^{-r}\right|_{x=0} = r(r+1)\cdots(r-k+1)$$

and we have

$$(1-x)^{-r} = \sum_{k=0}^{\infty} \frac{1}{k!} \frac{d^k}{dx} \left.(1-x)^{-r}\right|_{x=0} (x-0)^i$$

$$= \sum_{k=0} \frac{r(r+1)\cdots(r+k-1)}{k!} x^k = \sum_{k=0}^{\infty} \binom{r+k-1}{k} x^k = \sum_{n=r}^{\infty} \binom{n-1}{r-1} x^{n-r}$$

using above equality

using above equality

$$\sum_{n=r}^{\infty} P_X(n) = \sum_{n=r}^{\infty} \binom{n-1}{r-1} p^r (1-p)^{n-r} = p^r \cdot (1-(1-p))^{-r} = 1$$

④ Poisson distribution

Def. We say a random variable X has Poisson distribution with parameter λ, denoted by $Poi(\lambda)$ if

$$P_X(K) = e^{-\lambda} \frac{\lambda^K}{K!} \qquad K \in \{0, 1, 2, \ldots\}$$

Important equalities, resulted by Taylor expansion around 0:

$$e^x = \sum_{K=0}^{\infty} \frac{x^k}{K!}$$

$$e^{-x} = \sum_{K=0}^{\infty} \frac{(-x)^k}{K!}$$

$$\Rightarrow e^x + e^{-x} = 2 \sum_{K=0}^{\infty} \frac{x^{2K}}{(2K)!}$$