STA2001 Probability and Statistics (I)

Lecture 6

Tianshi Chen

The Chinese University of Hong Kong, Shenzhen

Review

Definition[Special mathematical expectation]

$$E[g(X)] = \sum_{x \in \overline{S}} g(x) f(x)$$

$$g(X) = \begin{cases} X \rightarrow \text{ Mean } = \mu \\ (X - E[X])^2 \rightarrow \text{ Variance } E[(X-\mu)^2] \\ X^r \rightarrow \text{ Moment} \\ e^{tX}, \text{ for } |t| < h, \rightarrow \text{ Mgf:} M(t) = \begin{cases} M(0) = 1 \\ M'(0) = E[X] \\ M''(0) = E[X^2] \end{cases} \end{cases}$$

Review

We are interested in the number of successes in n Bernoulli trials.

Definition[Binomial distribution]

A RV X is said to have a binomial distribution with n Bernoulli trials and the probability of success p, if the range space $\overline{S} = \{0, 1, \cdots, n\}$ and the pmf $f(x)$ is in the form of

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \cdots, n.$$

We can simply denote it by $X \sim b(n, p)$.

$X \sim b\ (n, p)$

Mgf of Binomial Distribution

Let $X \sim b(n, p)$. Then by definition,

$$M(t) = E[e^{tX}] = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x (1-p)^{n-x}$$

$$= [(1-p) + pe^t]^n \qquad -\infty < t < \infty$$

From the expansion of

$M(t) = [(1-p)+pe^t]^n$

$$(a+b)^n = \sum_{x=0}^{n} \binom{n}{x} a^x b^{n-x} \text{ with } a = pe^t, \quad b = 1-p$$

Mgf of Binomial Distribution

Question

What is the use of mgf?

Mgf of Binomial Distribution

Question

What is the use of mgf?

proof

$$M'(t) = n[(1-p)+pe^t]^{n-1}pe^t \Rightarrow M'(0) = E[X] = np$$

$$M''(t) = n(n-1)[(1-p)+pe^t]^{n-2}p^2e^{2t} + n[(1-p)+pe^t]^{n-1}pe^t$$

$$M''(0) = E[X^2] = n(n-1)p^2 + np$$

$$Var[X] = E[X^2] - (E[X])^2 = n^2p^2 - np^2 + np - n^2p^2 = np(1-p)$$

By the way, when $n = 1$ in $b(n,p)$, the binomial distribution reduces to Bernoulli distribution denoted by $b(1,p)$.

cdf of Binomial Distribution

$$F(x) = P(X \leq x) = \sum_{y \in \{X \leq x\}} f(y) = \sum_{y=0}^{[x]} \binom{n}{y} p^y (1-p)^{n-y},$$

where $x \in (-\infty, \infty)$ and $[x]$ is the largest integer $\leq x$.

Example 3

A kind of chicken are raised for laying eggs. Let $p = 0.5$ be the probability that the newly hacked chick is a female. Assuming independence, let X be the number of female chicken out of 10 newly hatched chicks selected at random.

$$P(X \leq 5)?$$

$$P(X = 6)?$$

$$P(X \geq 6)?$$

Example 3

Then $X \sim b(10, 0.5)$

$$P(X \leq 5) = \sum_{x=0}^{5} \binom{10}{x} 0.5^x 0.5^{5-x}$$

$$P(X = 6) = \binom{10}{6} 0.5^6 0.5^4 = P(X \leq 6) - P(X \leq 5)$$

$$P(X \geq 6) = 1 - P(X \leq 5)$$

Section 2.5 Negative Binomial Distribution

Negative Binomial Distribution

Description: We are interested in the number of Bernoulli trials

until exactly r successes occur, where r is a fixed positive integer.

负二项分布

$f(k;r,p) = \binom{k+r-1}{r-1} p^r (1-p)^k$

某事件在 $r+k$ 次实验中

出现第 r 次

Negative Binomial Distribution

Description: We are interested in the number of Bernoulli trials

until exactly r successes occur, where r is a fixed positive integer.

Define a RV X to denote the trial number at which the rth success

is observed. Then X has the range $\overline{S} = \{r, r+1, \cdots\}$.

Let $f(x)$ denote the pmf of X. Then recall $f(x) = P(X = x)$

Negative Binomial Distribution

第x次实验成功r次

$$f(x) = P(\{\text{at the } x\text{th trial, the } r\text{th success is observed}\})$$

$$= P(\{\underbrace{\text{for the first } x-1 \text{ trials, } r-1 \text{ success have been observed}}_{A}\}$$

$$\cap \{\underbrace{\text{at the } x\text{th trial, the outcome is a success}}_{B}\})$$

$$= P(A \cap B) = P(A)P(B)(\text{because } A \text{ and } B \text{ are independent})$$

=(第x-1次成功r-1次)·p

有一些书里的负二项分布的公式定义可能和这里的有一些小区别。最常见的变化就是：

X是实验总次数，得到r个失败的尝试。不仅仅是成功的次数。因此，实验总次数等于失败数加成功数，这个不同于这里定义的X。[3]

为了把公式换这种定义进行转换，把k用k-r代替，并且从均值、中位数，或者众数中减去r。为了将按本节定义的负二项分布的公式转换成本文里的公式，需要用k+r代替k，并且在均值，中位数，众数中加上r。

$$f(k;r,p) \equiv Pr(X=k) = \binom{k-1}{k-r} p^{k-r}(1-p)^r, k=r,r+1,r+2...$$

这个可能比上面的版本看起来更像二项分布，注意二项分布的参数是按顺序减少的：最后一个失败必然在最后发生，所以其它的事件有更少的可利用的位置，在计算顺序可能性时。

注意这里的负二项分布的定义没有推广到正实数r。

P表示失败的概率，不是成功的。为了把公式进行转换，每个地方用1-p代替p。X定义为失败次数，而不是成功的，这里的定义X为失败的，但P是成功的，和前面X表示成功但P表示失败概率的情况用同样的公式。但是失败和成功的描述是一致的，并且和前面的进行替换。

这两个替代公式可能会同时使用，比如X表示总次数，P表示失败次数。

负二项回归，分布是在均值m项里就定义了，并且和线性回归或者其它的一般线性回归的解释变量相关。概率密度函数变为

$$Pr(X=k) = \left(\frac{r}{r+m}\right)^r \frac{(k+r)!}{k!r!} \left(\frac{m}{r+m}\right)^k, k=0,1,2...$$

$$\begin{aligned} M(t) &= \sum_{x=r}^{\infty} e^{tx} \binom{x-1}{r-1} p^r (1-p)^{x-r} \\ &= (pe^t)^r \sum_{x=r}^{\infty} \binom{x-1}{r-1} [(1-p)e^t]^{x-r} \\ &= \frac{(pe^t)^r}{[1-(1-p)e^t]^r}, \qquad \text{where } (1-p)e^t < 1 \end{aligned}$$

(or, equivalently, when $t < -\ln(1-p)$). Thus,

$$\begin{aligned} M'(t) &= (pe^t)^r(-r)[1-(1-p)e^t]^{-r-1}[-(1-p)e^t] \\ &\quad + r(pe^t)^{r-1}(pe^t)[1-(1-p)e^t]^{-r} \\ &= r(pe^t)^r[1-(1-p)e^t]^{-r-1} \end{aligned}$$

and

$$\begin{aligned} M''(t) &= r(pe^t)^r(-r-1)[1-(1-p)e^t]^{-r-2}[-(1-p)e^t] \\ &\quad + r^2(pe^t)^{r-1}(pe^t)[1-(1-p)e^t]^{-r-1}. \end{aligned}$$

Accordingly,

$$M'(0) = rp^r p^{-r-1} = rp^{-1}$$

and

$$\begin{aligned} M''(0) &= r(r+1)p^r p^{-r-2}(1-p) + r^2 p^r p^{-r-1} \\ &= rp^{-2}[(1-p)(r+1)+rp] = rp^{-2}(r+1-p). \end{aligned}$$

Hence, we have

$$\mu = \frac{r}{p} \qquad \text{and} \qquad \sigma^2 = \frac{r(r+1-p)}{p^2} - \frac{r^2}{p^2} = \frac{r(1-p)}{p^2}.$$

Even these calculations are a little messy, so a somewhat easier way is given in Exercises 2.5-5 and 2.5-6.

Negative Binomial Distribution

$$
\begin{aligned}
f(x) &= P(\{\text{at the } x\text{th trial, the } r\text{th success is observed}\}) \\
&= P(\underbrace{\{\text{for the first } x-1 \text{ trials, } r-1 \text{ success have been observed}\}}_{A} \\
&\quad \cap \underbrace{\{\text{at the } x\text{th trial, the outcome is a success}\}}_{B}) \\
&= P(A \cap B) = P(A)P(B) (\text{because } A \text{ and } B \text{ are independent})
\end{aligned}
$$

$$
P(A) = \binom{x-1}{r-1} p^{r-1}(1-p)^{x-r}, \quad P(B) = p
$$

Therefore

$P(A)=\binom{x-1}{r-1}p^{r-1}(1-p)^{x-r}$ $P(B)=p.$

$$
f(x) = \binom{x-1}{r-1} p^{r}(1-p)^{x-r}, \quad x = r, r+1, \cdots
$$

$f(x)=\binom{x-1}{r-1}p^{r}(1-p)^{x-r}$

Negative Binomial Distribution

Definition[Negative Binomial Distribution]

A RV X is said to have a negative binomial distribution with the probability of success p and the number of successes r we are interested in, if the range $\overline{S} = \{r, r+1, \cdots\}$ and the pmf $f(x)$ is in the form of

$$f(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \quad x = r, r+1, \cdots .$$

This distribution get its name due to the negative binomial series

$$(1-w)^{-r} = \sum_{x=r}^{\infty} \binom{x-1}{r-1} w^{x-r}$$

$$(1-W)^{-r} = \sum_{X=r}^{\infty} \binom{X-1}{r-1} W^{X-r}$$

Geometric Distribution

Definition[Geometric Distribution]

A RV X is said to have a geometric distribution with the probability of success p, if the range $\overline{S} = \{1, 2, \cdots\}$ and the pmf $f(x)$ is in the form of 当Y=1的时候

$$f(x) = p(1-p)^{x-1}, \quad x = 1, 2, \cdots.$$

For a positive integer k,

$$P(X > k) = \sum_{x=k+1}^{\infty} p(1-p)^{x-1} = \frac{(1-p)^k p}{1-(1-p)} = (1-p)^k$$

$$P(X \leq k) = \sum_{x=1}^{k} p(1-p)^{x-1} = 1 - P(X > k) = 1 - (1-p)^k$$

Example 1, page 83

Biology students are checking eye color of fruit flies. For each fly,

$$P(\text{white}) = \frac{1}{4}, \quad P(\text{red}) = \frac{3}{4}.$$

Assume the observations are independent Bernoulli trials.

To observe 1 white fly, what's the probability one has to check

at least 4 flies?

at most 4 flies?

4 flies?

Example 1, page 83

We define X to be the number of fruit flies one has to check until the first white-eye fly is observed.

Then X has the geometric distribution with probability of success $1/4$. So the probability one has to check

at least 4 flies? $\longrightarrow P(X \geq 4) = P(X > 3) = (1 - \frac{1}{4})^3 = (\frac{3}{4})^3$

at most 4 flies? $\longrightarrow P(X \leq 4) = 1 - (1 - \frac{1}{4})^4$

4 flies? $\longrightarrow P(X = 4) = \frac{1}{4} \cdot (\frac{3}{4})^3$

Mathematical Expectations of Negative Binomial Distribution

Mean and Variance 负2项分布

$$\text{Mean}: E[X] = \frac{r}{p}$$

$r=1$ $u=\frac{1}{p}$ $\sigma^2=\frac{1-p}{p^2}$

$$\text{Variance}: Var[X] = E[X^2] - (E[X])^2 = \frac{r(1-p)}{p^2}$$

can be calculated by using the mgf

$$\text{Mgf}: M(t) = E[e^{tX}] = \frac{(pe^t)^r}{[1-(1-p)e^t]^r}, \text{ for } (1-p)e^t < 1$$

which can be obtained by using the negative binomial series

$$(1-w)^{-r} = \sum_{x=r}^{\infty} \binom{x-1}{r-1} w^{x-r}$$

Section 2.6 Poisson Distribution

Motivation

Description: There are experiments that result in counting the number of times that particular events occur within a given period or for a given physical object:

- the number of flaws in a 100 feet long wire.
- the number of customers that arrive at a ticket window between 7:00-8:00 pm.

Counting such events can be seen as observations of a RV associated with an approximate Poisson process (APP).

Approximate Poisson Process (APP)

Definition[Approximate Poisson Process (APP)]

Let the number of occurrences of some event in a given continuous interval be counted. Then we have an APP with parameter $\lambda > 0$ if

非重叠区间

(a) The number of occurrences in non-overlapping subintervals are independent. 独立

(b) The probability of exactly one occurrence in a sufficiently short subinterval of length h is approximately $\underline{\lambda h}$. $P = \lambda h$

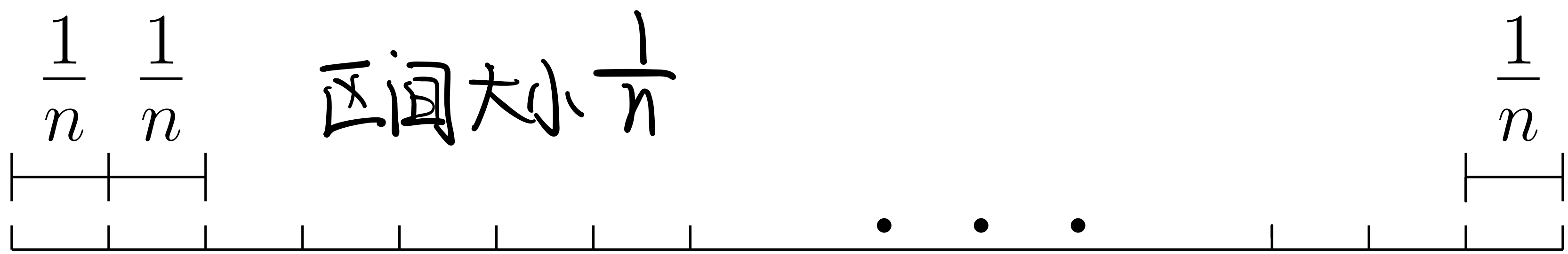
(c) The probability of two or more occurrences in a 足够小 sufficiently short subinterval is essentially $\underline{0}$.

Poisson Distribution

Consider a random experiment described by APP. Let X denote the number of occurrences in **an interval with length 1**. We aim to find an approximation for $f(x) = P(X = x)$ with $x = 0, 1, 2, \cdots$.

To this goal,

1. Partition the unit interval into n equally spaced subintervals.



2. If n is sufficiently large ($n >> x$), $P(X = x)$ can be approximated by the probability that exactly x of these n subintervals each has one occurrence.

$n >> x$

Poisson Distribution

2.1 By condition (c), the probability of two or more occurrences in any sufficiently short subinterval is 0. [n Bernoulli experiments.]

$\lambda \cdot \frac{1}{n}$

2.2 By condition (b), the probability of one occurrence in any subinterval (with length $\frac{1}{n}$) is approximately $\lambda\frac{1}{n}$. [Same probability of success $\lambda\frac{1}{n}$.]

2.3 By condition (a), the n Bernoulli experiments are independent. [n Bernoulli trials with probability of success $\lambda\frac{1}{n}$.]

Therefore occurrence and nonoccurrence in the n

subintervals are n Bernoulli trials with probability of success $\frac{\lambda}{n}$

$p = \frac{\lambda}{n}$

Poisson Distribution

3. Therefore, $P(X = x)$ can be approximated by the probability of x successes for $b(n, p = \frac{\lambda}{n})$

$$\frac{n!}{x!(n-x)!}(\frac{\lambda}{n})^x(1-\frac{\lambda}{n})^{n-x}$$

$b(n, \frac{\lambda}{n})$,

4. Let $n \to \infty$. Then

$$\lim_{n\to\infty} \frac{n!}{x!(n-x)!}(\frac{\lambda}{n})^x(1-\frac{\lambda}{n})^{n-x}$$

$$= \lim_{n\to\infty} \frac{n!}{(n-x)!n^x} \cdot \frac{\lambda^x}{x!}(1-\frac{\lambda}{n})^n(1-\frac{\lambda}{n})^{-x}$$

Poisson Distribution

$$\lim_{n\to\infty} \frac{n!}{x!(n-x)!}\left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n-x}$$

$$= \lim_{n\to\infty} \frac{n!}{x!(n-x)!}\left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n-x}$$

$$= \frac{n!}{(n-x)!\,n^x} \cdot \frac{\lambda^x}{x!} \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-x}$$

Noting

Noted that x is fixed.

$$\lim_{n\to\infty} \frac{n!}{(n-x)!n^x} = 1$$

$$\lim_{n\to\infty} (1-\frac{\lambda}{n})^n = e^{-\lambda},$$

$$\lim_{n\to\infty} (1-\frac{\lambda}{n})^{-x} = 1$$

We have

$$P(X=x) = \lim_{n\to\infty} \frac{n!}{(n-x)!n^x} \cdot \frac{\lambda^x}{x!}(1-\frac{\lambda}{n})^n(1-\frac{\lambda}{n})^{-x} = \frac{\lambda^x e^{-\lambda}}{x!}$$

$$P(X=x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Poisson Distribution

It can be verified

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, , \cdots$$

is a well-defined pmf.

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Definition[Poisson Distribution]

A RV X is said to have a Poisson distribution with the parameter λ, if the range $\overline{S} = \{0, 1, \cdots, \}$ and the pmf $f(x)$ is in the form of

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, , \cdots .$$

We can simply denote it by $X \sim \text{Poisson}(\lambda)$.

Poisson Distribution

It can be verified

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, , \cdots$$

is a well-defined pmf.

Definition[Poisson Distribution]

A RV X is said to have a Poisson distribution with the parameter λ, if the range $\overline{S} = \{0, 1, \cdots, \}$ and the pmf $f(x)$ is in the form of

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, , \cdots .$$

APP

We can simply denote it by $X \sim \text{Poisson}(\lambda)$.

Question

What's the implication of λ?

Mean and Variance

The mgf of a Poisson distributed RV X is

$$M(t) = E[e^{tX}] = \sum_{x=0}^{\infty} e^{tx} \frac{\lambda^x e^{-\lambda}}{x!}$$

$\sum_{x=0}^{\infty} e^{tx} \cdot \frac{\lambda^x e^{-\lambda}}{x!}$

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

$$= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} = e^{-\lambda} \cdot e^{\lambda e^t} = e^{\lambda(e^t - 1)}$$

$e^{\lambda e^t}$ 的 Taylor expansion

$$M'(t) = \lambda e^t e^{\lambda(e^t-1)} \Rightarrow M'(0) = \lambda$$

$$M''(t) = \lambda e^t e^{\lambda(e^t-1)} + \lambda^2 e^{2t} e^{\lambda(e^t-1)} \Rightarrow M''(0) = \lambda + \lambda^2 = E[X^2]$$

$$E[X] = M'(0) = \lambda$$

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

λ is the mean and variance of $X \sim$ Poisson(λ): the average number of occurrences in **the unit interval!**

avg occurrences in unit interval

Example 1, page 91

Question

In SZ, telephone calls to 110 come on the average of 2 calls every 3 minutes. If one models with APP, what's the probability of 5 or more calls arrive in a 9-minute period?

Example 1, page 91

Question

In SZ, telephone calls to 110 come on the average of 2 calls every 3 minutes. If one models with APP, what's the probability of 5 or more calls arrive in a 9-minute period?

Let X denote the number of calls in a 9-minute period, which is **the unit interval** for the problem. Then $X \sim \text{Poisson}(\lambda)$.

We need to determine λ.

Example 1, page 91

Question

In SZ, telephone calls to 110 come on the average of 2 calls every 3 minutes. If one models with APP, what's the probability of 5 or more calls arrive in a 9-minute period?

Let X denote the number of calls in a 9-minute period, which is **the unit interval** for the problem. Then $X \sim \text{Poisson}(\lambda)$.

We need to determine λ. $E[X] = 6 = \lambda$

$f(x) = \frac{6^x e^{-6}}{x!}$

$$E[X] = 6 = \lambda \Longrightarrow f(x) = \frac{6^x e^{-6}}{x!}$$

Therefore,

$$P(X \geq 5) = 1 - P(X \leq 4) = 1 - \sum_{x=0}^{4} \frac{6^x e^{-6}}{x!}$$