Mathematical Statistics 1

Ch.2 Discrete Distributions

Jungsoon Choi

jungsoonchoi@hanyang.ac.kr

Ch.2.6 The Poisson Distribution

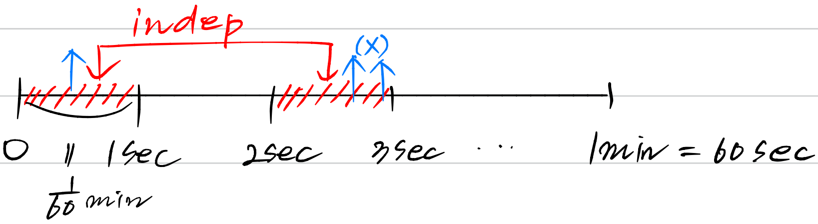
6.1 Approximate Poisson process

Let the number of occurrences of some event in a given continuous interval be counted. Then we have an approximate Poisson process with parameter $\lambda > 0$ if the following conditions are satisfied:

- The number of occurrences in non-overlapping subintervals are independent.
- The probability of exactly one occurrence in a sufficiently short subinterval of length h is approximately λh.
- The probability of two or more occurrences in a sufficiently short subinterval is essentially zero.

$X \sim$ Approximate Poissoon process with "λ"
(where λ means the averge counts in given time)

Let's consider 1 min as the time unit (1 min = 60 sec
1 sec = $\frac{1}{60}$ min)



① # of outcomes in [0,1] (sec) is indep of
# of outcomes in [2,3] (sec)

② P(one outcome appears in short sub-interval
of h $(=\frac{1}{60}$ min$)$) $\approx \lambda h = \frac{1}{60}\lambda$

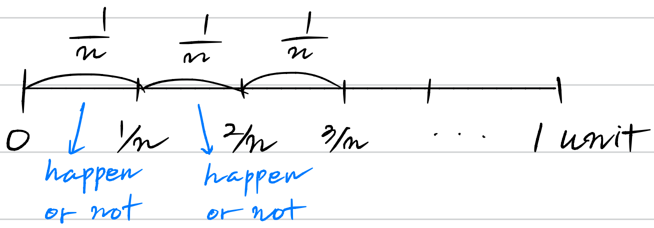
③ P(outcomes over two appear in short sub-interval) ≈ 0.

X = # of outcomes in time interval
$\Rightarrow f_X(x) = P(X=x)$

Derivation of Poisson process

X = # of outcomes in time interval "1"
Goal : We shold like to find an approximation for $P(X=x)$

1) Partition the time unit interval into large "n" sub-intervals of equal length "$1/n$" $= h$



$P(X=x) = P(x \text{ outcomes in "1" time unit})$
$\quad = P(x \text{ outcomes in "n" sub-intervals})$

ii) By approx. Poisson process (b),
$P(\text{exactly one outcome appears in the sub-interval})$
$\approx \lambda h = \frac{\lambda}{n}$

By approx. Poisson process (c),
two situations - outcome appears or not in the sub-interval

$\Rightarrow$ Sequence of "n" Bernoulli trials with $p = \frac{\lambda}{n}$ (prob. of occurence)

single outcome happens.

$P(X=x) \approx \binom{n}{x} p^x (1-p)^{n-x} = \binom{n}{x}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x}$

As n increases without bound,
$P(X=x) = \lim_{x\to\infty} \binom{n}{x}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x} =$

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

$$= \lim_{n\to\infty} \underbrace{\frac{n(n-1)\cdots(n-x-1)}{n^x}}_{\downarrow \, 1} \frac{\lambda^x}{x!} \underbrace{\left(1-\frac{\lambda}{n}\right)^n}_{\downarrow \, e^{-\lambda}} \underbrace{\left(1-\frac{\lambda}{n}\right)^{-x}}_{\downarrow \, 1} = \frac{\lambda^x e^{-\lambda}}{x!}$$

prob of Poisson distribution

X = # of outcomes in a given time interval

$$f_X(x) = P(X=x) = \frac{\lambda^x \cdot e^{-\lambda}}{x!} \qquad x = 0, 1, 2, \cdots$$

$$f_X(x) = P(X=x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!} \qquad x = 0, 1, 2, \cdots$$

$X \sim Poi(\lambda)$

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

6.2 Poisson Distribution, $\mathrm{Poi}(\lambda)$

- Let r.v X be the number of occurrences in a given time interval.
- *pmf*

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

- *Hint:* $\sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{\lambda}$ (Taylor series expansion)
- $P(X = x)$: dpois(x,lambda) in R
- $P(X \leq x)$: ppois(x,lambda) in R

- Mean: $E(X) = \lambda$
- Variance: $\mathrm{Var}(X) = \lambda$
- mgf: $M(t) = \exp\left\{\lambda(e^t - 1)\right\}$
- λ: average events within a given time interval (unit)

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

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

$$\overset{y=x-1}{=} e^{-\lambda} \sum_{y=0}^{\infty} \frac{\lambda^{y+1}}{y!} = e^{-\lambda} \cdot \lambda \cdot \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} = e^{-\lambda} \cdot \lambda \cdot e^{\lambda} = \lambda$$

$$E(X(X-1)) = \sum_{x=0}^{\infty} x(x-1) f_X(x) = \sum_{x=2}^{\infty} x(x-1) \frac{e^{-\lambda} \cdot \lambda^x}{x!} = e^{-\lambda} \sum_{x=2}^{\infty} \frac{\lambda^x}{(x-2)!}$$

$$\overset{y=x-2}{=} e^{-\lambda} \cdot \sum_{y=0}^{\infty} \frac{\lambda^2 \cdot \lambda^y}{y!} = e^{-\lambda} \cdot \lambda^2 \cdot e^{\lambda} = \lambda^2$$

$$Var(X) = E(X(X-1)) + E(X) - \{E(X)\}^2$$
$$= \lambda^2 + \lambda - \lambda^2 = \lambda.$$

mgf.

1) $\boxed{\begin{array}{l} E(X) = \lambda \\ Var(X) = \lambda \end{array}}$ $\Rightarrow$ λ : average outcomes in time interval.

Data1). $\underbrace{0.\ 1.\ 2.\ 3.\ 0.\ 2.\ 3\ \cdots}_{\text{average } = \lambda = 1.5}$ $\approx$ variance. $\sim$ Poi

Data2) 20. 10. 15. 0. 5.

$\Rightarrow$ average $<$ variance $\Rightarrow$ $\left\{\begin{array}{l} E(X) = \lambda \\ Var(X) = \lambda \end{array}\right.$ not satisfy

unit을 줄이면 $E(X) \approx Var(X)$. poi is not good distrib

time.

Data : 0 0 0 0 0 1 2 3 0 0

$\sim$ Poi(?). not generall poi

$\to$ Zero-inflated Poi

0-truncated Poi $x = 1, 2, 3, \cdots$

- Mean

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

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

- Variance

$$E[X(X-1)] = \sum_{x=0}^{\infty} x(x-1) \cdot f(x) = e^{-\lambda} \sum_{x=2}^{\infty} \frac{\lambda^x}{(x-2)!}$$

$$= \lambda^2 \cdot e^{-\lambda} \sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} = \lambda^2 \cdot e^{-\lambda} \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} = \lambda^2 \cdot e^{-\lambda} \cdot e^{\lambda} = \lambda^2$$

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

- *mgf*

$$M(t) = E\left(e^{tX}\right) = \sum_{x=0}^{\infty} e^{tx} f(x) = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} = \exp\left\{\lambda(e^t - 1)\right\}$$

$$M^{(1)}(0) = \frac{d}{dt} M(t)\bigg|_{t=0} = \frac{d}{dt} \exp\left\{\lambda(e^t - 1)\right\}\bigg|_{t=0} = \lambda e^t \exp\left\{\lambda(e^t - 1)\right\}\bigg|_{t=0} = \lambda$$

$$M^{(2)}(0) = \frac{d^2}{dt^2} M(t)\bigg|_{t=0} = \frac{d^2}{dt^2} \exp\left\{\lambda(e^t - 1)\right\}\bigg|_{t=0}$$

$$= \left(\lambda e^t\right)^2 \exp\left\{\lambda(e^t - 1)\right\} + \lambda e^t \exp\left\{\lambda(e^t - 1)\right\}\bigg|_{t=0} = \lambda^2 + \lambda$$

$$\mathrm{Var}(X) = M^{(2)}(0) - \left\{M^{(1)}(0)\right\}^2 = \lambda$$

Example 2.6-1

Let X have a Poisson distribution with a mean of $\lambda = 5$. Compute $P(X \leq 6)$, $P(X > 5)$, and $P(X = 6)$.

시험 X.

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

$$= \frac{e^{-5} \cdot 5^{6}}{6!}$$

Example 2.6-4

In a large city, telephone calls to 911 come on the average of two every 3 minutes. If one assumes an approximate Poisson process, what is the probability of five or more calls arriving in a 9-minute period?

2 / 3min.   5↑ / 9 min.

averiage. 2. calls / 3 min

" $2\times 3=6$. calls / 9 min

X = # of calls in 9 min.

$\sim \mathrm{Poi}(\lambda = 6)$

$P(X \geq 5)$.

6.3 Poisson approximation

- The Poisson r.v may be used as an approximation for a binomial r.v. with parameter (n, p) when n is large and p is small.
- $X \sim B(n, p)$ when $\lambda = np \ll \infty$, $n \to \infty$, and p is small.

$$
\begin{aligned}
f(x) = \binom{n}{x} p^x (1-p)^{n-x} &= \frac{n!}{x!(n-x)!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \\
&= \frac{n(n-1)\cdots(n-x+1)}{n^x} \frac{\lambda^x}{x!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x} \\
&\to \frac{e^{-\lambda}\lambda^x}{x!}, \qquad \text{since } \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}.
\end{aligned}
$$

$X \sim Bin(n, p)$

$f_X(x) = P(X=x) = \binom{n}{x} p^x (1-p)^{n-x}$

$n = 100, \quad p = 0.01$

$P(X=2) = \binom{100}{2} (0.01)^2 \cdot (0.99)^{98}$

$n \uparrow \infty$
$p \downarrow$ small $\Rightarrow \dfrac{e^{-\lambda} \cdot \lambda^x}{x!}$

$p = \frac{\lambda}{n}$ $\quad \lambda = np = 1$

λ : 계산결과가 1일뿐 반드시 1일 필요 X.

$P(X=2) \approx \dfrac{e^{-1} \cdot 1^2}{2!} = \dfrac{e^{-1}}{2} = \dfrac{1}{2e}$

① $n \geq 20$ & $p \leq 0.05$

② $n \geq 100$ & $p \leq 0.01$

Example 2.6-5

A manufacturer of Christmas tree light bulbs knows that 2% of its bulbs are defective. Assuming independence, we have a binomial distribution with parameters $p = 0.02$ and $n = 100$. Find the approximate probability that a box of 100 of these bulbs contains at most three defective bulbs.

$X \sim f_X(x) = P(X=x)$: pmf

$F_X(x) = P(X \le x)$: cdf

$E[g(x)] = \sum_{\forall x} g(x) f_X(x)$

$\hookrightarrow E(X) = \mu_X$

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

$M_X(t) = E[e^{tX}]$

Ber(p)

✓ Bin(n,p)

✓ NB(r,p) - 2 r.v.

Geo(p)

✓ Poi(λ)