Discrete Distributions

(1) Binomial (2) Negative Binomial (3) Geometric
(4) Poisson (5) Hypergeometric

1 Binomial Distribution

$$P(X=x) = \begin{cases} \binom{n}{x} p^x q^{n-x} & ; 0<p<1,\ q=1-p \\ & \ \ x=0(1)n \\ 0 & ; \text{o.w.} \end{cases}$$

$$\mu_{[r]}' = E(X)_r = (n)_r p^r$$

$$\mu_1' = E(X) = np$$

$$\mu_2 = \text{variance} = npq$$

\# Skewness $= \sqrt{\beta_1} = \dfrac{\mu_3}{\mu_2^{3/2}} = \dfrac{1-2p}{\sqrt{npq}}$

+vely skewed if $p < \frac{1}{2}$

-vely " " $p > \frac{1}{2}$

symmetric " $p = \frac{1}{2}$.

\# Kurtosis $= \gamma_2 = \beta_2 - 3 = \dfrac{\mu_4}{\mu_2^2} - 3$

$= \dfrac{1-6pq}{npq}$

leptokurtic if $pq < \frac{1}{6}$

mesokurtic " $pq = \frac{1}{6}$

platykurtic " $pq > \frac{1}{6}$

mean > variance

$V(x) \le \frac{n}{4}$ "=" holds when $p = q = \frac{1}{2}$

$\frac{x}{n} = f$ proportion of success.

$E(f) = p$

$V(f) = \frac{pq}{n}$

$V(f) \le \frac{\frac{n}{4}}{n^2} = \frac{1}{4n}$

$Cov\left(\frac{x}{n}, \frac{n-x}{n}\right) = -V\left(\frac{x}{n}\right) = \frac{-pq}{n}$

$P(X=x) = \frac{n-x+1}{x\,q}\,p\;P(X=x-1)$

let $P(X=x) = p_x$

$\frac{p_1}{p_0} \ge \frac{p_2}{p_1} \ge \cdots \ge \frac{p_n}{p_{n-1}}$

i.e. $\frac{P(X=x)}{P(X=x-1)}$ is a $\downarrow$ func^n of x.

PGF = $(q+zp)^n$, MGF = $(q+pe^t)^n$

① When $(n+1)p$ is an integer

then mode = $(n+1)p$ or $(n+1)p-1$

② When $(n+1)p$ is not an integer

mode = $[(n+1)p]$

If $X \sim Bin(2s, \frac{1}{2})$, $\frac{1}{2\sqrt{s}} < P(X=s) \le \frac{1}{\sqrt{2s+1}}$

$P(X=k) < \frac{1}{k-np}$ when $k > np$

If $np = \lambda$, then $\frac{\lambda^x}{x!}\left(1-\frac{x}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^n \le Bin(x; n, p) \le \frac{\lambda^x}{x!}\left(1-\frac{\lambda}{n}\right)^{n-x}$

Mean Deviation about mean.

$$MD_\mu = 2(n-k)\, p\, P(X=k) \simeq \sqrt{\frac{2npq}{\pi}}$$

$$\left[k = [np] \right]$$

$F(x) = I_q(n-x, x+1)$

$$= P(Z \le q) \qquad Z \sim Beta(n-x, x+1)$$

$$I_q(n-x, x+1) = \frac{1}{B(n-x, x+1)} \int_0^q z^{n-x-1}(1-z)^x \, dz$$

Negative Binomial Distribution

$$P(X=x) = \binom{x+r-1}{x} p^r q^x \quad , \; x=0,1,2,\ldots$$

$$E(X)_k = (r+k-1)_k \left(\frac{q}{p}\right)^k$$

$$\text{Mean} = \mu_1' = \frac{rq}{p}$$

$$\text{variance} = \mu_2 = \frac{rq}{p^2}$$

Mean < Variance

$$\frac{P(X=x)}{P(X=x-1)} = \frac{r+x-1}{x} q \quad ; \; x=1,2,\ldots$$

\# When $(r-1)\frac{q}{p} = k$ is an integer then mode $= (r-1)\frac{q}{p} - 1$ and $(r-1)\frac{q}{p}$.

When $\left[(r-1)\frac{q}{p}\right] = k$ and $(r-1)\frac{q}{p}$ is not an integer then mode $= \left[(r-1)\frac{q}{p}\right]$

\# $MD_\mu = 2\frac{k+1}{p} P(X=k+1) \qquad k=[np].$

\# $P(X \le k) = I_p(r, k+1)$

\# $P(X=x) = \binom{-r}{x} p^r (-q)^x$

$$= \binom{-r}{x} (-P)^x Q^{-(r+x)}$$

where $p=\frac{1}{Q}$, $q=\frac{P}{Q}$, $Q-P=1$ $\qquad Q=\frac{1}{p}$, $P=\frac{q}{p}$

$$M_X(t) = (Q - Pe^t)^{-r} \qquad P_X(t) = \left(\frac{1}{p} - \frac{q}{p}t\right)^{-r}$$

$$= \left(\frac{1}{p} - \frac{q}{p}e^t\right)^{-r}$$

X = number of failures preceding rth success in a sequence of Bernoulli trials with success probability p.

$$P(X \le k) = P(X + r \le k + r) = P(N \le k + r) = P(Z \ge r)$$

N = number of trials required to get rth success

Z = number of successout of $(k+r)$ trials.

$Z \sim Bin(k+r, p)$

$\displaystyle\sum_{k=0}^{b-1} \binom{a+k-1}{k} q^k = \sum_{k=0}^{b-1} \binom{a+b-1}{k} q^{b-1-k} p^k$

Let $X \sim Bin(n, p)$ and $Y \sim NB(r, p)$ denotes the number of trials required to get rth success in a sequence of independent Bernoulli trials then.

$$F_X(r-1) = 1 - F_Y(n)$$

SOURCE: GITHUB.COM/SOURAVSTAT

$\boxed{3}$

Geometric Distribution

$$P(X=x) = \begin{cases} pq^x, & x=0,1,2,\dots \text{ (Model I)} \\ pq^{x-1}, & x=1,2,3,\dots \text{ (Model II)} \end{cases}$$

$0<p<1$, $q=1-p$

Considering Model I,

$$E(x) = \frac{q}{p}$$

$$V(x) = \frac{q}{p^2}$$

$$\sum_{x=1}^{\infty} x q^{x-1} = \cancel{p}\,\frac{1}{(1-q)^2}$$

$$\Rightarrow \frac{d}{dq}\left[\sum_{x=1}^{\infty} x q^{x-1}\right] = \frac{d}{dx}\frac{1}{(1-q)^2}$$

$$\Rightarrow \sum_{x=2}^{\infty} x(x-1)\, q^{x-2} = \frac{2}{(1-q)^3}$$

Mean & Variance

PGF, $P_X(t) = \left(\frac{1}{p} - \frac{q}{p}t\right)^{-1}$

$$P(1+t) = p\left[1-q(1+t)\right]^{-1}$$

$$= \sum_{x=0}^{\infty}\left(\frac{qt}{p}\right)^x$$

$$\mu'_{[x]} = x!\left(\frac{q}{p}\right)^x$$

MGF. $M_x(t) = \left(\frac{1}{p} - \frac{q}{p}e^t\right)^{-1}$

Loss of Memory Property.

$x \sim Geo(p)$

Then $P(x \le t+k \mid x \ge t) = P(x \le k)$

This property characterizes ~~Geo~~ Geometric Distn

$P(X > i+j) = P(X > i)\,P(X \ge j)$

If x_1, x_2 are iid following $G(p)$. Then the conditional distribution $x_1 \mid x_1 + x_2$ is uniform.

$P(X_1 = x_1 \mid X_1 + X_2 = x) = \frac{1}{x+1},\ x_1 = 0, 1, 2, \ldots, x$

[For Model I]

For Model II, $P(X_1 = x_1 \mid X_1 + X_2 = x) = \frac{1}{x-1}\ \ I_{x_1 = 1, 2, \ldots, x-1}$

4

Poisson Distribution

$$P(X=x) = e^{-\lambda}\frac{\lambda^x}{x!}, \quad x=0,1,2,\cdots\infty;\ \lambda>0$$

$$E(X)_r = \lambda^r$$

$$E(X) = V(X) = \lambda.$$

$$\text{Skewness} = \gamma_1 = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\lambda}{\lambda^{3/2}} > 0$$

$\Rightarrow$ Poisson Dist^n is +vely skewed.

$$\text{Kurtosis} = \gamma_2 = \frac{\mu_4}{\mu_2^2} - 3 = \frac{3\lambda^2+\lambda}{\lambda^2} - 3 = \frac{1}{\lambda} > 0$$

$\Rightarrow$ Poisson Dist^n is leptokurtic.

MGF, $M_X(t) = e^{\lambda(e^t-1)}$

PGF, $P_X(t) = e^{\lambda(t-1)}$

$P(X=x) = \frac{\lambda}{x}P(X=x-1), \quad x=1,2,3,\cdots$

$MD_\mu(x) = E|X-\mu|$

$$= 2\lambda\left(\frac{e^{-\lambda}\lambda^{x_0}}{x_0!}\right), \quad x_0 = [\lambda]$$

$$\rightarrow \sqrt{\frac{2\lambda}{\pi}}$$

$X \sim P(\lambda=1) \Rightarrow MD_{\mu}(x) = \frac{2}{e} \times s.d.$

$E\left(\frac{1}{1+x}\right) = \frac{1}{\lambda}\left(1-e^{-\lambda}\right)$; $x \sim P(\lambda)$

$E(X \cdot g(X)) = \lambda E[g(X+1)]$ if $X \sim P(\lambda)$

If $X \sim P(\lambda)$, then, $E(X^r) = \lambda E(X+1)^{r-1}$

$\mu_{r+1} = \lambda \sum_{i=1}^{r} \binom{r}{i} \mu_{r-i}$

$P(X \leq k) = 1 - \frac{\int_0^{\lambda} e^{-t} t^k \, dt}{\Gamma k+1}$

If $X \sim P(\lambda)$, then, $P(X \geq n) < \frac{\lambda^n}{n!}$

If $x_1, x_2, \ldots x_n$ are independently distributed poisson variates and $x_i \sim P(\lambda_i)$ $\forall$ $i = 1(1)n$. then $S_n = \sum_1^n x_i \sim P(\sum_1^n \lambda_i)$

Suppose x_1 and x_2 are independently distributed poisson variates $\ni$ $x_1 \sim P(\lambda_1)$, $x_2 \sim P(\lambda_2)$, then,

$$x_1 \mid x_1 + x_2 = k \sim Bin\left(k, \frac{\lambda_1}{\lambda_1+\lambda_2}\right)$$

If $X \sim P(\lambda)$ and $Y \mid X = x \sim Bin(x, p)$, then $Y \sim P(\lambda p)$

Hypergeometric Distribution

$$P(X=x) = \frac{\binom{Np}{x}\binom{Nq}{n-x}}{\binom{N}{n}} = \binom{n}{x}\frac{(Np)_x\,(Nq)_{n-x}}{(N)_n}$$

$x = 0, 1, 2, \dots, n$

$0<p<1$

\# $Hyp(N; n, p) \to Bin(n, p)$ when $N \to \infty$.

\# $E(X)_r = \dfrac{(n)_r\,(Np)_r}{(N)_r}$

\# Mean : $\mu_1' = np$

\# Variance $= npq\,\dfrac{N-n}{N-1} \to npq$ as $N \to \infty$

\# If $\dfrac{(n+1)(Np+1)}{N+2} = k$ is an integer then there are two modes at $k-1$ and at k.

If k isn't an integer then the mode is unique at $[k]$.

\# $MD_\mu = 2(n-k)\left(p - \frac{k}{N}\right) P(X=k)$

Here $k = [np]$

\# $V(X) \le \frac{n}{4}$.