University of Akron, Dept. of Statistics

3470:651 Probability and Statistics

Common Discrete Distributions

Textbook: Casella and Berger 2ed. (2013)

(September 21, 2016)

Contents

3.1 Binomial

Analogy: X is Number of heads in n tosses of a coin. $P(H) = p$ for each toss.

$X \sim Bin(n, p)$

$$\begin{aligned}
\text{pmf}: \quad p(x) &= P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad \text{for } x = 0, 1, 2, \ldots n. \\
\text{CDF}: \quad F(x) &= P(X \leq x) = \sum_{k=0}^{x} p(x) \\
\text{mean}: \quad E(X) &= np \\
\text{var}: \quad V(X) &= np(1-p) \\
\text{MGF}: \quad M(t) &= \left[e^t p + (1-p)\right]^n
\end{aligned}$$

Called Bernoulli Distribution if $r = 1$.

```
dbinom(2, n,  p)       #pmf at x=2
pbinom(2, n,  p)       #CDF at x=2
pbinom(.5, n, p)       #Inv CDF at q=.5
rbinom(1000, n, p)     # random sample of size 1000
```

3.2 Negative Binomial

Analogy: Number of **tails** until you get r heads.

$X \sim NegBin(r, p)$

$$
\begin{aligned}
\text{pmf}: \quad p(x) &= \binom{x+r-1}{r-1}(1-p)^x p^r \quad \text{for } x = 0, 1, 2, \ldots \\
\text{CDF}: \quad F(x) &= P(X \leq x) = \sum_{k=0}^{x} p(x) \\
\text{mean}: \quad E(X) &= \frac{r(1-p)}{p} \\
\text{var}: \quad V(X) &= \frac{r(1-p)}{p^2} \\
\text{MGF}: \quad M(t) &= [\frac{p}{1-(1-p)e^t}]^r
\end{aligned}
$$

Called Geometric Distribution if $r = 1$.

```
dnbinom(2, r,  p)       #pmf at x=2
pnbinom(2, r,  p)       #CDF at x=2
pnbinom(.5, r,  p)      #Inv CDF at q=.5
rnbinom(1000, r, p)     # random sample of size 1000
```

Negative Binomial (Flips ver.)

Analogy: Number of **flips** until you get r heads.

$X \sim NegBin(r, p)$

$$\begin{aligned}
\text{pmf}: \quad p(x) &= \binom{x-1}{r-1} p^r (1-p)^{x-r} \quad \text{for } x = r, r+1, r+2, \ldots \\
\text{CDF}: \quad F(x) &= P(X \le x) = \sum_{k=0}^{x} p(x) \\
\text{mean}: \quad E(X) &= \frac{r}{p} \\
\text{var}: \quad V(X) &= \frac{r(1-p)}{p^2} \\
\text{MGF}: \quad M(t) &= [\frac{pe^t}{1-(1-p)e^t}]^r
\end{aligned}$$

Called Geometric Distribution if $r = 1$.

```
dnbinom( 10-r, r,  p)       #pmf at x=10  (In R, X=# of tails)
pnbinom( 10-r, r,  p)       #CDF at x=10
pnbinom(.5, r,  p)      #Inv CDF at q=.5
rnbinom(1000, r, p)     # random sample of size 1000
```

3.3 Hypergeometic

Analogy: N balls in an urn, of which m are red. Pick n at once. $X=$ number of red balls.

$X \sim HG(n, m, N)$

$$\begin{aligned}
\text{pmf}: \quad p(x) &= p(X=x) = \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}} \qquad ** \\
\text{CDF}: \quad F(x) &= P(X \leq x) = \sum_{k=0}^{x} p(x) \\
\text{mean}: \quad E(X) &= np \\
\text{var}: \quad V(X) &= \left(\frac{N-n}{N-1}\right) np(1-p) \\
\text{MGF}: \quad M(t) &= DoesNotExist
\end{aligned}$$

where $p = m/N$.

** for $\max(0, n-N+m) \leq x \leq \min(n, m)$, and 0 otherwise.

```
dhyper(2, m, N-m, n)        #pmf at x=2
phyper(2, m, N-m, n)        #CDF at x=2
phyper(.5, m, N-m, n)       #Inv CDF at q=.5
rhyper(1000, m, N-m, n)      # random sample of size 1000
```

3.4 Poisson

Analogy: events with rate λ per unit time.

$X \sim Poi(n, m, N)$

$$\begin{aligned}
\text{pmf}: \quad p(x) &= p(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \text{ for } x = 0, 1, 2, \ldots \\
\text{CDF}: \quad F(x) &= P(X \leq x) = \sum_{k=0}^{x} p(x) \\
\text{mean}: \quad E(X) &= \lambda \\
\text{var}: \quad V(X) &= \lambda \\
\text{MGF}: \quad M(t) &= \exp\{\lambda(e^t - 1)\}
\end{aligned}$$

```
dpois(2, lambda)       #pmf at x=2
ppois(2, lambda)       #CDF at x=2
qpois(.5, lambda)      #Inv CDF at q=.5
rpois(1000, lambda)     # random sample of size 1000
```

Notes

- ..

R-code

$X \sim \text{Bin}(n = 10, p = .4)$

```
X = rbinom(1000, 10, .4)
plot(X)
hist(X)

dbinom(x=2, 10, .4)
pbinom(x=2, 10, .4)

t = seq(0,10)
plot( t, dbinom(t,10,.4) )
plot( t, pbinom(t,10,.4),type='s' )
```

$X \sim \text{NB}(r = 5, p = .4)$ [X=num of Failures]

```
X = rnbinom(1000, 5, .4)
plot(X)
hist(X)

dnbinom(x=2, 5, .4)
pnbinom(x=2, 5, .4)

t = seq(0,10)
plot( t, dnbinom(t,5,.4) )
plot( t, pnbinom(t,5,.4),type='s' )
```

$X \sim \mathrm{HG}(n = 15, m = 10, N = 30)$

```
X = rhyper(1000, 10,20,15)
plot(X)
hist(X)

dhyper(x=2, 10,20,15)
phyper(x=2, 10,20,15)

t = seq(0,15)
plot( t, dhyper(t,10,20,15) )
plot( t, phyper(t,10,20,15),type='s' )
```

Multinomial Distribution

Analogy: Throw a die n times that has k sides, $\{C_1, \dots, C_k\}$, with probability $\{p_1, \dots, p_k\}$. The outcome $\{X_1, \dots, X_k\}$ represents the frequency of each outcome.

- pmf:

$$p(x_1, \dots, x_k) = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k}$$

Example: Multinomial

- Suppose there are Three buckets, A, B, C. If you throw a ball,

$$P(A) = .3, \quad P(B) = .2, \quad P(C) = .1.$$

- If you throw a ball 20 times, what is the probability that

$$P(\text{ Exactly two A, three B, five C })$$

3.5 Detail Calculations

3.5.1 Binomial

Binomial Coefficient and Binomial Expantion

$$(a+b)^n = \sum_{x=0}^{n} \binom{n}{x} a^x b^{n-x}$$

So if you expand

$$(x+y)^{25} = x^{25} + \cdots + \quad \binom{25}{6} x^6 y^{19} \quad + \cdots + y^{25}$$

Check if the sum of pmf is 1

$$\begin{aligned}
\sum_{x=0}^{n} p(x) &= \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} \quad \text{(use binomial expantion formula)} \\
&= \Big[p + (1-p)\Big]^n = 1
\end{aligned}$$

Expectation and Variance via pmf

$$
\begin{aligned}
E(X) &= \sum_{x=0}^{n} x p(x) \\
&= \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x}
\end{aligned}
$$

The first term $x = 0$ is zero.

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

Then, we use identity $x\binom{n}{x} = n\binom{n-1}{x-1}$ (see blow). So the sum becomes

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

taking np out of the sum,

$$
= np \sum_{x=1}^{n} \binom{n-1}{x-1} p^{x-1} (1-p)^{n-1-(x-1)}
$$

Leting $k = x - 1$,

$$= \quad np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

The sum is 1 because it's just a pmf of Bin(n-1,p)

$$= \quad np$$

Identitiy

$$\begin{aligned} x\binom{n}{x} &= x\frac{n!}{x!(n-x)!} = \frac{n!}{(x-1)!(n-x)!} \\ &= \frac{n(n-1)!}{(x-1)!(n-1-(x-1))!} = n\binom{n-1}{x-1} \end{aligned}$$

MGF

$$
\begin{aligned}
M(t) = E(e^{tX}) &= \sum_{x=0}^{n} \binom{n}{x} e^{tx} p^x (1-p)^{n-x} \\
&= \sum_{x=0}^{n} \binom{n}{x} (e^t p)^x (1-p)^{n-x} \qquad \text{(use binomial expantion formula)} \\
&= \Big[e^t p + (1-p)\Big]^n
\end{aligned}
$$

Expectation via MGF

$$
\begin{aligned}
M(t) &= \left[e^t p + (1-p)\right]^n \\
M'(t) &= n\left[e^t p + (1-p)\right]^{n-1} e^t p \qquad\qquad M'(0) = E(X) = np \\
M''(t) &= n(n-1)\left[e^t p + (1-p)\right]^{n-2} e^{2t}p^2 + n\left[e^t p + (1-p)\right]^{n-1} e^t p \\
M''(0) &= E(X^2) = n(n-1)p^2 + np = n^2p^2 - np^2 + np
\end{aligned}
$$

Thus the variance is $V(X) = E(X^2) - E(X)^2 = np(1-p)$.

Expectation via independent Bernoulli

3.5.2 Negative Binomial

3.5.3 Hypergeometric

Binomial vs Hypergeometric

- Suppose you have populatio of N subject, of which m are defective. Let

$$X = \text{[number of defectives in sample]} .$$

- Sample with replacement.

- Sample without replacement.

3.5.4 Poisson

Check sum of pmf is 1

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

MGF

$$\begin{aligned} M(t) = E(e^{tX}) &= \sum_{x=0}^{\infty} e^{tx} e^{-\lambda} \lambda^x / x! \\ &= e^{-\lambda} \sum_{x=0}^{\infty} (\lambda e^t)^x / x! \\ &= e^{-\lambda} \exp\{\lambda e^t\} \\ &= \exp\{\lambda(e^t - 1)\} \end{aligned}$$

Poisson as a limit of Binomial

Poisson distribution is the limit of binomial distribution when $n \to \infty$, $p \to 0$, in such a way that $np \to \lambda$.

Starting from Binomial pmf and replacing $p = \lambda/n$,

$$\begin{aligned}
p_X(x) = P(X = x) &= \binom{n}{x} p^x (1-p)^{n-x} \\
&= \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \\
&= \frac{1}{x!} \frac{n!}{(n-x)!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x} \\
&= \frac{1}{x!} \left(\frac{n!}{(n-x)! n^x}\right) \lambda^x \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x}
\end{aligned}$$

If we take the lim,

$$\begin{aligned}
\lim_{n\to\infty} p_X(x) = P(X = x) &= \lim_{n\to\infty} \frac{1}{x!} \left(\frac{n!}{(n-x)! n^x}\right) \lambda^x \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x} \\
&= \frac{\lambda^x}{x!} e^{-\lambda}
\end{aligned}$$

Sum of Poisson is Poisson

- mgf for poisson

$$
\begin{aligned}
M_{X_1}(t) = E(e^{tX}) &= \sum_{x=0}^{\infty} e^{tx} \frac{e^{-\lambda}\lambda^x}{x!} \\
&= \sum_{x=0}^{\infty} \frac{e^{-\lambda}(e^t\lambda)^x}{x!} \\
&= \frac{e^{-\lambda}\sum_{x=0}^{\infty}(e^t\lambda)^x}{x!} \\
&= e^{-\lambda}e^{(\lambda e^t)} = e^{\lambda(e^t-1)}
\end{aligned}
$$

- If $X_1, X_2 \sim$Poi(λ) and independent, since

$$M_{X_1+X_2}(t) = M_{X_1}(t)M_{X_1}(t) = e^{\lambda(e^t-1)}e^{\lambda(e^t-1)} = e^{2\lambda(e^t-1)},$$

we see taht $X_1 + X_2 \sim$Poi(2λ).

Poisson process

- Let $N(t)$ denote the number of events before time t

$$P(N(t_2) - N(t_1) = x) = \frac{e^{-\lambda(t_2-t_1)}\left[\lambda\left(t_2 - t_1\right)\right]^x}{x!}.$$

- Assume independence over disjoint time interval.

- Then waiting time between events will be iid Exponential with mean $1/\lambda$.