

Discrete distributions

Lecture 10

Binomial distribution

x_1, \dots, x_n i.i.d. $\text{Bern}(p)$ r.v., part of the data is

$$\begin{aligned} P(\bar{X}_1 = x_1, \dots, \bar{X}_n = x_n) \\ = p^{\sum x_i} q^{n - \sum x_i} \end{aligned}$$

One-parameter exponential $T = \sum \bar{X}_i$, complete sufficient statistic.

$T \in \text{Bin}(n, p)$.

$f(p)$ U-estimable, mean

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

for some estimator δ , i.e. if we let be a polynomial of degree $\leq n$.

If f is a polynomial of degree $\leq n$, then

(A') We can use (1) to identify

①

an estimator δ . This means solving for δ in (1).

or alternatively

(B) Use conditioning.

We can then for any $m \leq n$, form

$$P(\bar{X}_1=1, \dots, \bar{X}_m=1) = p^m$$

T sufficient, and therefore

$$\mathcal{J}(T) = P(\bar{X}_1=1, \dots, \bar{X}_m=1 | T)$$

$$= E(1\{\bar{X}_1=1, \dots, \bar{X}_m=1\} | T)$$

is an UMVU estimator of its expectation

i.e. of $E(1\{\bar{X}_1=1, \dots, \bar{X}_m=1\}) = P(\bar{X}_1=1, \dots, \bar{X}_m=1)$

$$= p^m.$$

But $\bar{T} = \sum_{i=1}^n \bar{X}_i$. Therefore if

$T = t < m$ then

$$\mathcal{J}(t) = 0$$

②

If $T = t \geq m$ then

$$\delta(t) = P(\bar{X}_1 = 1, \dots, \bar{X}_m = 1 \mid \sum_{i=1}^n \bar{X}_i = t)$$

$$= \frac{P(\bar{X}_1 = 1, \dots, \bar{X}_m = 1 \wedge \sum \bar{X}_i = t)}{P(\sum \bar{X}_i = t)}$$

$$\frac{\cancel{P(m \text{ success in the first } m \text{ trials and } t-m \text{ successes in the remaining } n-m \text{ trials})}}{P(T=t)}$$

$$= \frac{p^m \binom{n-m}{t-m} p^{t-m} q^{n-t}}{\binom{n}{t} p^t q^{n-t}}$$

$$= \frac{t(t-1)\dots(t-m+1)}{n(n-1)\dots(n-m+1)}$$

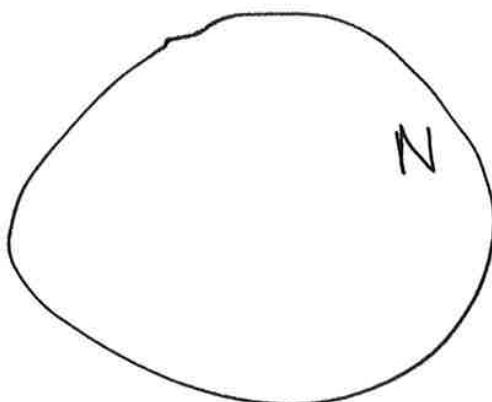
This is the UMVU estimator of

$$p^m.$$

③

#

Capture-recapture problem



N total number of fish (or a certain species)

First catch a number of fish, of that species, and tag them and let loose. Suppose K fish have been tagged.

Now draw a fixed number n of fish from the pond (drawing with replacement), and note that X (rv.) number have been caught before

$$X \in \text{Bin}(n, p) = \text{Bin}(n, \frac{K}{N})$$

Know K from before. If we can estimate p , then we can estimate N

(9)

$$\cancel{p = \frac{K}{N}} \Leftrightarrow N = \frac{K}{p}$$

We need an unbiased estimator of $\frac{1}{p}$.
This will give us an unbiased estimator
of N number of fish in the pond.
But, recall: there is no ~~an~~ unbiased
estimator of $\frac{1}{p}$ when using
binomial sampling.

Inverse binomial sampling

Intuitively one needs more data to
estimate $\frac{1}{p}$ when p is close to
zero; small change in p will large
change in $\frac{1}{p}$.

Sample by taking more observations,
then you would with binomial sampling
for smaller p .

Inverse sampling

Means that continue to sample until we get a specified number of (tripped fish) successes, number m .

Assume one ~~weed~~ needs $y+m$ trials to get m ~~successes~~ successes (Y failures). Then

$$(1) \quad P(Y=y) = \binom{m+y-1}{m-1} p^m q^y, \quad y=0, 1, 2, \dots$$



$$E(Y) = \frac{mq}{p}, \quad \text{Var}(Y) = \frac{mq}{p^2}.$$

(We have done this)

Let's define estimator

$$\delta(Y) = \frac{Y+m}{m}$$

Then

$$\begin{aligned} E(\delta(Y)) &= \frac{E(Y+m)}{m} = \frac{\frac{mq}{p} + m}{m} = \\ &= \frac{\frac{mq+mp}{p}}{mp} = \underline{\underline{\frac{1}{p}}} \end{aligned}$$

Also (1) is a one-parameter exp-family
with Y complete sufficient statistic (Problem 36)

Therefore $\delta(Y)$ is UMVU

#

Power series distributions

This is a one-parameter exponential family of form

$$P(\bar{X}=x) = a(x)\theta^x \frac{1}{C(\theta)}, \quad x=0, 1, 2, \dots$$

$(\theta > 0)$.

⑦

We see that since

$$I = \sum_{x=0}^{\infty} P(\bar{X}=x) = \sum_{x=0}^{\infty} a(x) \theta^x \frac{1}{C(\theta)}$$

\Leftrightarrow

$$C(\theta) = \sum_{x=0}^{\infty} a(x) \theta^x$$

which means that $a(x)$ are the coefficients in the power series expansion of $C(s)$.

Ex:

$$a(x) = \begin{cases} \binom{n}{x} & x = 0, 1, \dots, n \\ 0 & \text{otherwise} \end{cases}$$

$$\theta = \frac{p}{q}$$

$$C(\theta) = (\theta + 1)^n$$

is a binomial distribution. #

Ex.

$$a(x) = \binom{m+x-1}{m-1}$$

$$\theta = p$$

$$C(\theta) = (1-\theta)^{-m}$$

negative binomial

(8)

#

Assume $a(x) > 0$ for all x . We want to find unbiased estimator δ of θ^r (r positive integer).

Solve for δ in

$$\sum_{x=0}^{\infty} \delta(x) a(x) \theta^x = C(\theta) \theta^r$$

$$x=0$$

$$\sum_{x=0}^{\infty} \delta(x) a(x) \theta^{x-r} = C(\theta)$$

$$x=0$$

$$\sum_{x=0}^{\infty} \delta(x) a(x-r) \theta^{x-r} \frac{a(x)}{a(x-r)} = C(\theta)$$

Since

$$C(\theta) = \sum_{x=0}^{\infty} a(x) \theta^x$$

therefore

$$\delta(x) = \begin{cases} 0 & x=0, 1, \dots, r-1 \\ \frac{a(x-r)}{a(x)} & x \geq r \end{cases}$$

An unbiased estimator of θ^r .

(9)

Lemme.

X_1, \dots, X_n are i.i.d. $P(X_i = x) = q(x)\theta^x / C(\theta)$

The distribution of $T = X_1 + \dots + X_n$

$$P(T=t) = \frac{A(t,n) \theta^t}{(C(\theta))^n}$$

with $A(t,n)$ the coefficient of θ^t
in the power series expansion of $(C(\theta))^n$

Proof.

We have

$$\begin{aligned} P(T=t) &= \sum_{\substack{x_1 + \dots + x_n = t}} \frac{\theta^{x_1} q(x_1)}{C(\theta)} \cdots \frac{\theta^{x_n} q(x_n)}{C(\theta)} \\ &= \theta^t \frac{1}{C(\theta)^n} A(t,n) \end{aligned}$$

where

$$A(t,n) = \sum_{\substack{x_1 + \dots + x_n = t}} q(x_1) \cdots q(x_n)$$

The last is clear since a pmf
has to sum to one.

10

Can we find the complete sufficient statistic?

Note that X_1, \dots, X_n i.i.d. with above power series distribution, have a joint density

$$P(X_1 = x_1, \dots, X_n = x_n) \\ = \left(\prod_{i=1}^n g(x_i) \right) \frac{\theta^{x_1 + \dots + x_n}}{C(\theta)^n}$$

($p(x) = g_\theta(\bar{x}) h(\bar{x})$), by factorization criterion $T = X_1 + \dots + X_n$ is sufficient.

Also T is complete, hence . . .