Review

- Last lecture - inference for a proportion, p

→ random sample from population

→ series of independent binary observations

$Z_i: \{ \begin{smallmatrix} 0 \\ 1 \end{smallmatrix} \quad Pr(Z_i = 1) = p$

likelihood function

maximize → maximum likelihood estimate

$$\frac{\partial \log \text{likelihood}}{\partial p} = 0$$

→ likelihood eqn
score eqn

$$I(p) = -\frac{\partial^2 \log \text{likelihood}}{\partial p^2}$$

$$\mathcal{I} = E(I(p)) = \text{Fisher information}$$

over data

$Var(\hat{p}) = 1/\mathcal{I}$ → s.e.

Binomial - large sample inference
- small sample exact

Multinomial - touched on

Poisson Distribution

discrete distⁿ $X = 0, 1, 2 \ldots$

- counts
- numbers of events in time and/or space

- limiting case of binomial

$\mu = E(X) = np$ (binomial) - hold fixed but let $n \to \infty$ $\left(p = \frac{\mu}{n} \to 0\right)$

$$P(X = x) \to \frac{\mu^x e^{-\mu}}{x!} \to \text{Poisson dist}^n$$

example - Binomial type assumptions of independence
- not clear what n is

ex. radiation expt
- typical pattern
  - unimodal
  - skewed positively

If μ is large (> 5) distⁿ looks ~~like~~ fairly normal

$\mu = E(X)$ - max. lik. est is $\bar{X} = 3.87$

for inference $s.e.(\bar{x}) = ?$

$\rightarrow$ theoretically $Var(X) = \mu$

$$Var(\bar{X}) = \mu/n \qquad \hat{\mu} = \bar{x}$$

$$s.e.(\hat{\mu}) = \sqrt{\bar{X}/n}$$

ex. $\sqrt{\frac{3.87}{2608}} = .0385$

95% Conf. Int - 3.792 - 3.943

Note on calculation - tabulated data

$$\bar{X} = \frac{\sum_{x=0}^{\infty} f_x \times x}{n} \qquad f_x = \text{freq of } x$$
$$n = \sum f_x$$

<u>Model checking</u> - check Poisson assumption

- discrete data - QQ plots don't work too well

Can calculate expected frequencies under Poisson assumption

expected # 0's $= \underset{=2608}{n} \times Pr(X=0) = \frac{\mu^0 e^{-\mu}}{0!}$

insert $\hat{\mu} = 3.87 \rightarrow 54.54$

Continuing for all x values

expected # 1's $= n \times \mu^1 e^{-\mu}/1! = 210.94$

Informally, fit seems good.

Can conduct a formal goodness-of-fit test, by calculating discrepancy measure (describes "distance" of obs'd from exp'd freq).

- Calculate P-value if we can assess the sampling distⁿ of measure under H_0: Poisson distⁿ

Multinomial theory yields

$$\chi^2 = \text{Pearson's stat} = \sum \frac{(\text{obs'd} - \text{exp'd})^2}{\text{exp'd}}$$

If we define finite # of categories → special case of multinomial

categories 0, 1, 2, . . . , (10 or more)

So for last category - obs'd = 16

expected = $n \times \text{Prob}(X \geq 10)$
$(1 - P(X=0) \ldots P(X=9))$

Expect = 17

Based on multinomial theory, if

E(obs'd) = exp'd - i.e. model holds

X^2 will follow a χ^2 distn

with d.f. depends on

k = # cells (k = 11)

p = # parameters estimated to calculate exp'd ($p = 1$ $\hat{\mu}$)

$df = k - p - 1$

ex. $X^2 = \frac{(57 - 54.54)^2}{54.54} + \cdots = 12.9$

P-value = Prob$(\chi^2_9 > 12.9) = .167$

- slight logical issue with goodness-of-fit
  - since models rarely exactly true, large enough sample will eventually detect it.
    - plot is more informative diagnostic

an alternative to X^2 - Pearson.

derived more directly from likelihood via likelihood ratio test

$G^2 = 2 \sum \text{obs'd} \times \log\left(\frac{\text{obs'd}}{\text{exp'd}}\right)$

$\doteq$ follows χ^2_{k-p-1} d.f.

ex . G^2 13.97 $\sim$ p-value = .125

So model appears satisfactory

Exact conf. interval for μ,
based on tail areas of
Poisson distⁿ, which conveniently
can be derived from χ^2 tails

→ X = total of observations $(\hat{\lambda} = X/n)$

$$\left(\chi^2_{\underbrace{(1-\alpha/2)}_{\text{upper tail areas}},\ \underbrace{2x}_{\text{d.f.}}},\ \chi^2_{\alpha/2,\ 2x+2} \right)$$

is an exact $(1-\alpha) \times 100\%$ Conf. Int
for μ.

(3.792, 3.944) — agrees to 2nd decimal

Inference for Rates

- often mortality
  disease, morbidity

Poisson distⁿ central.

example. leukemia

- very large n
- small risk - of disease
  - rare disease

→ Person-years - measure of
  # individuals
  × length of time
  followed.

- or may be Σ times
  with different times for
  different individuals

If we observe X = # cases of
disease

X will be approx. Poisson

$$\mu = E(X) = \underset{\text{rate}}{\lambda} \times \underset{\text{person-year of f.u.}}{T}$$

f.u. follow-up

In data table. 4 Poisson observation

$$\hat{\mu}_i = x_i \qquad \hat{\lambda}_i = x_i / T$$

So. - rate for < 1

$X = 2 \to$ exact CI for μ

$$\frac{.242 - 7.22}{\div 2.25}$$ from χ^2 fla

Conf for λ = rate per 100K

$\hat{\lambda} = \frac{2}{\text{pop}^n} \times 100\text{k}$

= .88

CI for λ .108 to 3.21