# M339J: April 14th, 2023.

## MLE: Binomial Distribution.

$X \sim \text{Binomial}(m, q)$ — $m$: $\boxed{\text{positive integer}}$, $q \in (0,1)$

In general, both $m$ and $q$ are unknown.

If you have complete, unmodified data: $x_1, x_2, \dots, x_n$

Most commonly $\boxed{m \text{ is known}}$

Then, the likelihood f'tion is

$$L(q) = \prod_{j=1}^{n} f_X(x_j; q) = \prod_{j=1}^{n} \binom{m}{x_j} q^{x_j} \cdot (1-q)^{m-x_j}$$

$$\propto \prod_{j=1}^{n} q^{x_j} \cdot \prod_{j=1}^{n} (1-q)^{m-x_j} = q^{\Sigma x_j} (1-q)^{m \cdot n - \Sigma x_j}$$

$\Rightarrow$ The loglikelihood f'tion:

$$\ell(q) = \text{const} + (\Sigma x_j) \ln(q) + (m \cdot n - \Sigma x_j) \cdot \ln(1-q)$$

$$\Rightarrow \ell'(q) = (\Sigma x_j) \cdot \frac{1}{q} + (m \cdot n - \Sigma x_j) \cdot \frac{1}{1-q} (-1) = 0$$

$$\frac{\Sigma x_j}{q} = \frac{m \cdot n - \Sigma x_j}{1-q}$$

$$(\Sigma x_j)(1-\cancel{q}) = q\,(m \cdot n - \cancel{\Sigma x_j})$$

$$\boxed{\hat{q}_{MLE} = \frac{\bar{x}}{m} = \frac{\text{total \# of successes}}{\text{total \# of observations}}}$$ ✓

**53.** You are given:

(i) The distribution of the number of claims per policy during a one-year period for 10,000 insurance policies is:

| Number of Claims per Policy | Number of Policies |
|---|---|
| 0 | 5000 |
| 1 | 5000 |
| 2 or more | 0 |

(ii) You fit a binomial model with parameters $m$ and $q$ using the method of maximum likelihood.

Calculate the maximum value of the loglikelihood function when $m = 2$.

(A) −10,397

(B) −7,781

(C) −7,750

(D) −6,931

(E) −6,730

$\longrightarrow$: $X \sim \text{Binomial}(m=2, q)$

$L(q) = \left(\binom{2}{0} q^0 (1-q)^2\right)^{5000} \cdot \left(\binom{2}{1} q^1 \cdot (1-q)^1\right)^{5000}$

$L(q) = 2^{5000}\, q^{5000} (1-q)^{15000}$

$\ell(q) = 5000 \cdot \ln(2) + 5000 \cdot \ln(q) + 15000 \ln(1-q)$

$\hat{q}_{MLE} = \dfrac{\text{total \# of successes}}{n \cdot m} = \dfrac{5000}{20000}$

$\boxed{\hat{q} = \frac{1}{4}}$

$\ell\left(\frac{1}{4}\right) = 5000 \cdot \ln(2) + 5000 \ln(1/4) + 15000 \ln(3/4) = \underline{-7780.97}$

# Mortality Laws.

**Example.** $X \sim \text{Exponential}(\theta)$

**Example.** $X \sim U(0, \theta)$ $\boxed{\theta > 0}$ : pdf : $f_X(x;\theta) = \frac{1}{\theta}$ $x \in (0, \theta)$

$\boxed{x_1, x_2, \dots, x_n \text{ data set}}$

Goal: MLE on $\theta$

Likelihood ftion:

$L(\theta) = \left(\frac{1}{\theta}\right)^n$ w/ $x_j \leq \theta$ for all $j = 1, \dots, n$,

i.e., $\boxed{\max(x_1, x_2, \dots, x_n) \leq \theta}$ ←

Note: $L(\theta)$ is strictly decreasing as a f'tion of $\theta$

⇒ It attains its maximum @ $\boxed{\max(x_1, x_2, \dots, x_n) = \hat{\theta}_{MLE}}$

**Def'n.** The hazard rate (also known as the force of mortality or the failure rate) of a r.v. $X$ is a function $h_X : I \to \mathbb{R}$ defined as

$$\boxed{h_X(x) = \frac{f_X(x)}{S_X(x)}} = \frac{-S_X'(x)}{S_X(x)} = -\frac{d}{dx}\left[\ln(S_X(x))\right]$$

for all $x \in I$ where $I$ is the subset of $\mathbb{R}$ where $f_X$ is well-defined and $S_X \neq 0$.

Q: How do you express the survival ftion in terms of $h_X$?

→: $h_X(x) = -\frac{d}{dx}\left[\ln(S_X(x))\right]$

$-h_X(x)\,dx = d\left[\ln(S_X(x))\right]$

$\int$ \

$-\int_0^b h_X(x)\,dx = \ln(S_X(b))$ (+ const)

$$\boxed{S_X(b) = e^{-\int_0^b h_X(x)\,dx}}$$

Other notation: $\lambda_X(x) \dots$ rates

$\boxed{\mu_x \dots \text{force of mortality}}$

**Example.** $X \sim U(0,\theta)$

$h_X(x) = ?$

$x \in (0,\theta)$ : $h_X(x) = \dfrac{f_X(x)}{S_X(x)} = \dfrac{\frac{1}{\theta}}{1-\frac{x}{\theta}} = \boxed{\dfrac{1}{\theta - x}}$

**Example.** $X \sim \text{Exponential}(\theta)$

**18.6** You are doing a mortality study of insureds between ages 70 and 90. Two specific lives contributed this data to the study:

| Life | Age at Entry | Age at Exit | Cause of exit |
|---|---|---|---|
| 1 | 70.0 | 90.0 | End of study |
| 2 | 70.0 | Between 89.0 and 90.0 | Death |

You assume mortality follows Gompertz law $\mu_x = B \times c^x$ and plan to use maximum likelihood estimation.

$L$ is the likelihood function associated with these two lives.

$L^*$ denotes the value of $L$ if the Gompertz parameters are $B = 0.000003$ and $c = 1.1$.

Calculate $L^*$.

(A) 0.0115

(B) 0.0131

(C) 0.0147

(D) 0.0163

(E) 0.0179

**18.7** You are doing a mortality study of insureds between ages 60 and 90. Two specific lives contributed this data to the study:

| Life | Age at Entry | Age at Exit | Cause of exit |
|---|---|---|---|
| 1 | 60.0 | 74.5 | Policy lapsed |
| 2 | 60.0 | 74.5 | Death |

You assume mortality follows Gompertz law $\mu_x = B \times c^x$ and plan to use maximum likelihood estimation.

$L$ is the log-likelihood function (using natural logs) associated with these two lives.

$L^*$ denotes the value of $L$ if the Gompertz parameters are $B = 0.000004$ and $c = 1.12$.

Calculate $L^*$.

(A) -4,67

(B) -4.53

(C) -4.39

(D) -4.25

(E) -4.11

**18.8** You are given the following seriatim data on survival times for a group of 12 lives. The superscript $+$ indicates a right-censored value.

$$25,\ 32^+,\ 35^+,\ 36,\ 40^+,\ 44,\ 48,\ 60,\ 62^+,\ 65,\ 67,\ 70^+$$

Calculate the standard deviation of the estimate of $S(50)$ using the Nelson-Aalen estimator.

(A) 0.1455

(B) 0.1519

(C) 0.1547

(D) 0.1621

(E) 0.1650

[Question on October 2022 FAM-L Exam]