Stat581 Assignment 6

Tianzhi Cao (tc796)

Abhishek Bhatt (ab2083)

Siva Harshini Dev Bonam (sdb202)

October 21, 2020

1 Bernoulli

Bernoulli's p.d.f. is:$f(x;p) = p^x(1-p)^{1-x}$. Estimate p, we get L_n:

$$L_n(p) = \prod_{i=1}^{n} p^{X_i}(1-p)^{1-X_i} = p^{\sum_{i=1}^{n} X_i}(1-p)^{n-\sum_{i=1}^{n} X_i}$$

Then we use optimize() function maximize that to get the p value.

2 Geometric

Geometric's p.d.f. is:$f(x;p) = p * (1-p)^x$. Estimate p, we get L_n:

$$L_n(p) = \prod_{i=1}^{n} p * (1-p)^{X_i} = p^n * (1-p)^{\sum_{i=1}^{n} X_i}$$

$$ln(L_n(p)) = n * ln(p) + \sum_{i=1}^{n} X_i * ln(1-p)$$

Then we use optimize() function maximize that to get the p value.

3 Normal

Normal's p.d.f. is:$f(x;\mu;\sigma) = (1/\sqrt{2*\pi*\sigma^2})*exp(-\frac{(x-\mu)^2}{2\sigma^2})$.

$$L_n(\mu;\sigma) = \prod_{i=1}^{n}(1/\sqrt{2*\pi*\sigma^2})*exp(-\frac{(x-\mu)^2}{2\sigma^2})$$

$$L_n(\mu;\sigma) = (2\pi\sigma^{-n})^{-n/2}*exp(-\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{2\sigma^2})$$

Take ln for L_n we get:

$$ln(L_n(\mu;\sigma)) = ln((2\pi\sigma^{-n})^{-n/2}) + (-\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{2\sigma^2})*ln(e)$$

$$\frac{(L_n(\mu;\sigma))}{\partial\mu} = \frac{\sum_{i=1}^{n}(X_i-\mu)}{\sigma^2}$$

Because we need maximum L_n, $\frac{(L_n(\mu;\sigma))}{\partial\mu}$ should equal to 0.

$$\frac{\sum_{i=1}^{n}(X_i-\mu)}{\sigma^2} = 0$$

$$\sum_{i=1}^{n}(X_i) = n*\mu$$

Then we took μ back to

$$ln(L_n(\mu;\sigma)) = ln((2\pi\sigma^{-n})^{-n/2}) + (-\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{2\sigma^2})*ln(e)$$

and use optimize() function maximize that to get the σ

4 Chisq

Chisq's p.d.f. is:$f(x;k) = \frac{x^{k/2-1}*e^{-x/2}}{2^{k/2}*\Gamma(k/2)}$.

$$L_n(k) = \prod_{i=1}^{n}\frac{x_i^{k/2-1}*e^{-x_i/2}}{2^{k/2}*\Gamma(k/2)}$$

$$ln(L_n(k)) = \sum_{i=1}^{n}ln(\frac{x_i^{k/2-1}*e^{-x_i/2}}{2^{k/2}*\Gamma(k/2)})$$

$$ln(L_n(k)) = (k/2-1)\sum_{i=1}^{n}ln(x_i) - 1/2\sum_{i=1}^{n}x_i - nln(\Gamma(k/2)) - (nk/2)ln(2)$$

and use optimize() function maximize that to get the k.

5 Multivariate Normal

If a multivariate normal distribution has K dimentions, it p.d.f. is: $f(x_1, ..., x_k; \mu; \sigma) = \frac{1}{(2\pi)^{K/2}} \frac{1}{cov(x)^{1/2}} e^{-1/2(x-\mu)^T cov(x)^{-1}(x-\mu)}$

$$L_n(\mu; \sigma) = \prod_{i=1}^{n} \frac{1}{(2\pi)^{K/2}} \frac{1}{cov(x)^{1/2}} e^{-1/2(x-\mu)^T cov(x)^{-1}(x-\mu)}$$

$$= \frac{1}{(2\pi)^{Kn/2}} \frac{1}{cov(x)^{n/2}} e^{-1/2 \sum_{i=1}^{n}(x_i-\mu)^T cov(x)^{-1}(x_i-\mu)}$$

$$ln(L_n(\mu; \sigma)) = -(Kn/2)ln(2*\pi) - (N/2)lncov(x) - (1/2)\sum_{i=1}^{n}(x_i-\mu)^T cov(x)^{-1}(x_i-\mu)$$

Taking derivative of μ

$$\frac{\partial ln(L_n)}{\partial \mu} = -2\sum_{i=1}^{n} cov(x)^{-1}x_i + 2ncov^{-1}\mu$$

When $\frac{\partial ln(L_n)}{\partial \mu} = 0$ we have max $ln(L_n)$.

$$0 = -2\sum_{i=1}^{n} cov(x)^{-1}x_i + 2ncov^{-1}\mu$$

$$\mu = (1/n)\sum_{i=1}^{n} x^n = \bar{x}$$

Then we took μ back to

$$ln(L_n(\mu; \sigma)) = -(Kn/2)ln(2*\pi) - (N/2)lncov(x) - (1/2)\sum_{i=1}^{n}(x_i-\mu)^T cov(x)^{-1}(x_i-\mu)$$

Taking derivative of (x)

$$\frac{\partial ln(L_n)}{\partial (x)} = (cov(x)^{-1} - cov(x)^{-1}\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^T cov(x)^{-1})^T$$

When $\frac{\partial ln(L_n)}{\partial (x)} = 0$ we have max $ln(L_n)$.

$$0 = cov(x)^{-1} - cov(x)^{-1}\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^T cov(x)^{-1})^T$$

$$cov(x) = 1/n\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})^T$$

6 Binomial

Probability Mass Function:

$$f(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}$$

Likelihood function:

$$L_n(p, n) = \prod_{i=1}^{t} \binom{n}{X_i} p^{X_i} (1-p)^{n-X_i}$$

$$= p^{\sum_{i=1}^{t} X_i} (1-p)^{tn - \sum_{i=1}^{t} X_i} \prod_{i=1}^{t} \binom{n}{X_i}$$

Log-likelihood function:

$$l_n(p, n) = ln(\prod_{i=1}^{t} \binom{n}{X_i}) + ln(p) \sum_{i=1}^{t} X_i + ln(1-p)(n - \sum_{i=1}^{n} X_i)$$

For a fixed n (number of independent trials/samples), the optimization is then,

$$argmax_p[ln(p) \sum_{i=1}^{t} X_i + ln(1-p)(tn - \sum_{i=1}^{t} X_i)]$$

We solve the optimization in R using optimize() function to get the corresponding value of p as the estimated parameter.

7 Exponential

Probability Density Function:

$$f(x; \lambda) = \lambda e^{-\lambda x}, x \geq 0$$

Likelihood function:

$$L_n(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda X_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} X_i}$$

Log-likelihood function:

$$l_n(\lambda) = nln(\lambda) - \lambda \sum_{i=1}^{n} X_i$$

The optimization is then,

$$argmax_{\lambda}[nln(\lambda) - \lambda \sum_{i=1}^{n} X_i]$$

We solve the optimization in R using optimize() function to get the corresponding value of λ as the estimated parameter.

8 Beta

Probability Density Function:

$$f(x; \alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}$$

where B(α, β) is called the Beta function.
Likelihood function:

$$L_n(\alpha, \beta) = \prod_{i=1}^{n} \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} = \frac{1}{B(\alpha, \beta)^n} \prod_{i=1}^{n} X_i^{\alpha-1}(1-X_i)^{\beta-1}$$

Log-likelihood function:

$$l_n(\alpha, \beta) = -nln(B(\alpha, \beta)) + \sum_{i=1}^{n} [(\alpha-1)ln(X_i) + (\beta-1)ln(1-X_i)]$$

We optimize the log likelihood of Beta distribution with Limited-memory BFGS optimization algorithm to estimate parameters α and β.

9 Multinomial

The likelihood which can be described as joint probability is

$$L(\vec{p}\,) = n! \prod_{i=1}^{m} \frac{p_i^{x_i}}{x_i!}$$

where $(p_1, \ldots, p_m)$ is the vector of probabilities with $\sum_{i=1}^{m} p_i = 1$ and $n = \sum_{i=1}^{m} x_i$ is the total number of independent trials/samples.

The log-likelihood is

$$\begin{aligned} l(\vec{p}\,) = \log L(\vec{p}\,) &= \log\left(n! \prod_{i=1}^{m} \frac{p_i^{x_i}}{x_i!}\right) && (1) \\ &= \log n! + \log \prod_{i=1}^{m} \frac{p_i^{x_i}}{x_i!} && (2) \\ &= \log n! + \sum_{i=1}^{m} \log \frac{p_i^{x_i}}{x_i!} && (3) \\ &= \log n! + \sum_{i=1}^{m} x_i \log p_i - \sum_{i=1}^{m} \log x_i! && (4) \end{aligned}$$

Posing a constraint $\sum_{i=1}^{m} p_i = 1$ with Lagrange multiplier

$$l'(\vec{p}\,, \lambda) = l(\vec{p}\,) + \lambda\left(1 - \sum_{i=1}^{m} p_i\right) \tag{5}$$

To find $\arg\max_{\vec{p}} L(\vec{p}, \lambda)$

$$\frac{\partial}{\partial p_i} l'(\vec{p}, \lambda) = \frac{\partial}{\partial p_i} l(\vec{p}) + \frac{\partial}{\partial p_i} \lambda \left(1 - \sum_{i=1}^{m} p_i\right) = 0 \tag{6}$$

$$\frac{\partial}{\partial p_i} \sum_{i=1}^{m} x_i \log p_i - \lambda \frac{\partial}{\partial p_i} \sum_{i=1}^{m} p_i = 0 \tag{7}$$

$$\frac{x_i}{p_i} - \lambda = 0 \tag{8}$$

$$p_i = \frac{x_i}{\lambda} \tag{9}$$

Since we have

$$p_i = \frac{x_i}{\lambda} \tag{10}$$

$$\sum_{i=1}^{m} p_i = \sum_{i=1}^{m} \frac{x_i}{\lambda} \tag{11}$$

$$1 = \frac{1}{\lambda} \sum_{i=1}^{m} x_i \tag{12}$$

$$\lambda = n \tag{13}$$

Therefore

$$\hat{p_i} = \frac{x_i}{n} \tag{14}$$

So we get estimated probability vector as

$$\hat{\vec{p}} = \left(\frac{x_1}{n}, ..., \frac{x_m}{n}\right) \tag{15}$$

10 Point Mass

The probability mass function for point mass distribution is $f(x) = 1$ in $x = a$ and 0 otherwise. This means f can have only two values, either 1 and 0, which implies that the $\mathcal{L}(\theta)$ which is the product of many fs, we can choose the MLE for point mass distribution to be mode among X_n.

11 Poisson

$$\mathcal{L}(\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = e^{-n\lambda} \frac{\lambda^{\sum_1^n x_i}}{\prod_{i=1}^n x_i!}$$

$$lnL(\lambda) = -n\lambda + \sum_1^n x_i ln(\lambda) - ln\left(\prod_{i=1}^n x_i!\right)$$

$$\frac{dlnL(\lambda)}{dp} = -n + \sum_1^n \frac{x_i}{\lambda}$$

$$\hat{\lambda} = \frac{\sum_{i=1}^n x_i}{n}$$

12 Uniform

$$\mathcal{L}_n(a,b) = \frac{1}{(b-a)^n}$$

if all $a \leq X_1, X_2, ..., X_n \leq b$ else $\mathcal{L} = 0$ To maximize, we need to minimize $b-a$, the boundary condition is to choose $\hat{a} = min_i(X_i)$ and $\hat{b} = max_i(X_i)$

13 Gamma

There are two parameters in the Gamma distribution - alpha and beta. Estimation of beta is -

$$\beta = mean(inputData)/alpha$$

So, first we need to approximate the alpha parameter. Alpha is approximated using the following equation

$$lnL(\alpha) - digamma(\alpha) = lnL(mean(inputData)) - mean(log(inputData))$$

There is no close formed solution for alpha, so we approximate it using the following approximation -

$$lnL(\alpha) - digamma(\alpha)\ (1/2 * \alpha)(1 + 1/(6 * \alpha + 1))$$

If we take

$$s = lnL(mean(inputData)) - mean(log(inputData))$$

Then, we approximate alpha using the following solution -

$$alpha = (3 - s + \sqrt{((s-3)**2 + 24*s))}/12*s$$

And then using this value of alpha we can easily estimate the beta value

14 Hypergeometric

Probability Mass Function:

$$f(x; N, K, n) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$$

if $x = \max(0, n - N + K), \ldots, \min(K, n)$, otherwise 0

where N is the population size, K is the number of success states in the population, n is the number of units drawn in each trial and x is the number of observed successes.

Therefore likelihood function of N given the observation y is

$$L(N \mid y) = \frac{\binom{K}{y}\binom{N-K}{n-y}}{\binom{N}{n}}, N \in \{1, 2, 3, \ldots\}$$

Because the domain for N is the nonnegative integers, we cannot use calculus. However, we can look at the ratio of the likelihood values for successive value of the total population.

$$\frac{L(N \mid y)}{L(N-1 \mid y)} = \frac{N-n}{N} \cdot \frac{N-K}{N-K-n+y}$$

Now studying when this ratio exceeds 1, equals 1 and is less than 1, we can find out when $L(Ny)$ reaches its maximum. MLE of N is the value of N at which the likelihood is maximized. $L(N|y) > L(N1|y)$ for $N < \lfloor nK/y \rfloor$ and $L(N|y)L(N1|y)$ for $N\lfloor nk/y \rfloor$. This gives the maximum likelihood estimator

$$\hat{N}(y) = \left\lfloor \frac{nK}{y} \right\rfloor$$