6.5 Activity II: Estimation of Parameters

- One of more hypothesized distribution
- We must specify the values of their parameters
- Many methods to estimate distribution parameters
  - Method of moments
  - Unbiased
  - Least squares
  - Maximum likelihood estimator (MLE)
• MLEs have several statistical properties
  • For most of the common distributions, the MLE is unique
  • Although MLEs need not be unbiased, in general, the asymptotic distribution of \( \theta \) has mean equal to \( \theta \).
  • MLEs are invariant
  • MLEs are asymptotically normal distributed
  • MLEs are strongly consistent ( \( \lim_{n \to \infty} \hat{\theta} = \theta \) (w.p. 1) )
Idea for MLEs:

- Have observed sample $X_1, X_2, ..., X_n$
- Came from some true (unknown) parameter(s) of the distribution form
- Pick the parameter(s) that make it most likely that you would get what you did get (or close to what you got in the continuous case)
- An optimization (mathematical-programming) problem, often messy
MLEs for Discrete Distributions

- Have hypothesized family with probability mass function $p_\theta (x)$
- Single (for now) unknown parameter $\theta$ to be estimated
- Define the likelihood function $L(\theta)$
  $$L(\theta) = p_\theta (X_1) p_\theta (X_2) \ldots p_\theta (X_n)$$
- Find the (legal) value of $\theta$ that maximizes $L(\theta)$
MLEs for Continuous Distributions

- Change “getting” above to “getting close to” for motivation (see Prob. 6.26)
- Wind up just replacing probability mass function $p_\theta$ by density $f_\theta$ and proceed the same way

$$L(\theta) = f_\theta (X_1) f_\theta (X_2) \ldots f_\theta (X_n)$$

- Find the (legal) value of $\theta$ that maximizes $L(\theta)$
Example of Continuous MLE: Interarrival-Time Data for Drive-Up Bank

Hypothesized exponential family: density function is

\[ f_\beta(x) = \begin{cases} \frac{1}{\beta} e^{-x/\beta} & \text{if } x > 0 \\ \text{Otherwise} \end{cases} \]

Likelihood function is

\[ L(\beta) = \left( \frac{1}{\beta} e^{-x_1/\beta} \right) \left( \frac{1}{\beta} e^{-x_2/\beta} \right) \ldots \left( \frac{1}{\beta} e^{-x_n/\beta} \right) = \beta^{-n} \exp \left( -\frac{1}{\beta} \sum_{i=1}^{n} X_i \right) \]

Want value of \( \beta \) that maximizes \( L(\beta) \) over all \( \beta > 0 \)

Equivalent (and easier) to maximize the log-likelihood function \( l(\beta) = \ln L(\beta) \) since \( \ln \) is a monotonically increasing function
In this case, \( l(\beta) = -n \ln \beta - \frac{1}{\beta} \sum_{i=1}^{n} X_i \), which can be maximized by simple differential calculus:

Set \( \frac{dl}{d\beta} = -\frac{n}{\beta} + \frac{1}{\beta^2} \sum_{i=1}^{n} X_i = 0 \) and solve for \( \beta = \frac{\sum_{i=1}^{n} X_i}{n} = \bar{X}(n) \)

Check second-order sufficient conditions for a maximizer:

\[ \frac{d^2l}{d\beta^2} = \frac{n}{\beta^2} - 2 \frac{2}{\beta^3} \sum_{i=1}^{n} X_i \], which is negative when \( \beta = \bar{X}(n) \) since the \( X_i \)'s are positive.

Thus, the MLE is \( \hat{\beta} = \bar{X}(n) = 0.399 \) from the observed sample of \( n = 219 \) points.
Example of Discrete MLE: Demand-Size Data from Inventory

Hypothesized geometric family: mass function is \( p_{p}(x) = p(1 - p)^{x} \) for \( x = 0, 1, 2, \ldots \)

Likelihood function is \( L(p) = p^{n} (1 - p)^{\sum_{i=1}^{n} X_{i}} \)

In this case, log-likelihood function is \( l(p) = n \ln p + \sum_{i=1}^{n} X_{i} \ln(1 - p) \), which can be maximized by simple differential calculus:

Set \( \frac{dl}{dp} = \frac{n}{p} - \frac{\sum_{i=1}^{n} X_{i}}{1 - p} = 0 \) and solve for \( p = \frac{1}{X(n) + 1} \)

Check second-order sufficient conditions for a maximizer:

\[
\frac{d^{2}l}{dp^{2}} = -\frac{n}{p^{2}} - \frac{\sum_{i=1}^{n} X_{i}}{(1 - p)^{2}}, \text{ which is negative for any valid } p
\]
So MLE is \( \hat{p} = \frac{1}{1.891 + 1} = 0.346 \) from the observed sample of \( n = 156 \) points.

Confidence interval for true \( p \):

\[
E\left( \frac{d^2 l}{dp^2} \right) = -\frac{n}{p^2} - \frac{\sum_{i=1}^{n} E(X_i)}{(1 - p)^2} = -\frac{n}{p^2} - \frac{n(1 - p)}{p} = -\frac{n}{p^2 (1 - p)}
\]

Thus, \( \delta(p) = p^2 (1 - p) \) and for large \( n \), an approximate 90% confidence interval for \( p \) is

\[
\hat{p} \pm 1.645 \sqrt{\frac{\hat{p}^2 (1 - \hat{p})}{n}}
\]

\[
0.346 \pm 1.645 \sqrt{\frac{0.346^2 (1 - 0.346)}{156}}
\]

\[
0.346 \pm 0.037
\]

\[ [0.309, 0.383] \]
6.6 Activity III: Determining How Representative the Fitted Distributions Are

- Have: Hypothesized family, have estimated parameters
- Question: Does the fitted distribution agree with the observed data?
- Approaches: Heuristic and formal statistical hypothesis tests
6.6.1 Heuristic Procedures

- **Density/Histogram Overplots and Frequency Comparisons**
  - *Continuous Data*
    - Density/histogram overplot:
    - Plot $\hat{b} f(x)$ over the histogram $h(x)$; look for similarities
Interarrival-time data for drive-up bank and fitted exponential:

\[
\hat{f}(x) = \begin{cases} 
2.506e^{-x/0.399} & \text{if } x \geq 0 \\
0 & \text{otherwise}
\end{cases}
\]

\[\hat{\beta} = 0.399\]
**Frequency comparison**

Histogram intervals interval $[b_{j-1}, b_j]$ for $j = 1, 2, ..., k$, each of width $\Delta b$

Let $h_j = \text{the observed proportion of data in } j\text{th interval}$

Let $r_j = \int_{b_{j-1}}^{b_j} \hat{f}(x) \, dx$, the *expected* proportion of data in $j$th interval if the fitted distribution is correct

Plot $h_j$ and $r_j$ together, look for similarities
• **Discrete Data**
  
  • **Frequency comparison**

  Let $h_j = \text{the observed proportion of data that are equal to the } j\text{th possible value } x_j$

  Let $r_j = \hat{p}(x_j)$, the expected proportion of the data equal to $x_j$ if the fitted probability mass function $\hat{p}$ is correct

  Plot $h_j$ and $r_j$ together, look for similarities

  Demand-size data for inventory and fitted geometric:
\[ \hat{p}(x) = \begin{cases} 
0.346(0.654)^x & \text{if } x = 0,1,2,\ldots \\
0 & \text{otherwise}
\end{cases} \]

\[ \hat{p} = 0.346 \]
**Distribution Function Differences Plots**

- The density/histogram overplot is comparison of *individual probabilities of* fitted distribution and observed *individual distribution*.
- Define a empirical distribution function $F_n(x)$
  
  $$F_n(x) = \frac{\text{number of } X_i \text{ 's } \leq x}{n} = \text{ proportion of data that are } x$$

- Could plot $\hat{F}(x)$ with $F_n(x)$ *and look for similarities, but it is harder to see such similarities for cumulative than for individual probabilities*
- Alternatively, plot $\hat{F}(x)$ $F_n(x)$ *- against the range of x values and look for closeness to a flat horizontal line at height 0*
Interarrival-time data for drive-up bank and fitted exponential:

Demand-size data for inventory and fitted geometric:
• **Probability Plots**
• Another class of ways to compare CDF of fitted distribution with an empirical directly from the data
• Sort data into increasing order: $X_1, X_2, \ldots, X_n$ (*called the order statistics of the data*)
• A reasonable estimate of the distribution function $F(x)$ of $i/n$.
• Empirical distribution function

$$\tilde{F}_n(X_{(i)}) = F_n(X_{(i)}) - \frac{0.5}{n} = \frac{i - 0.5}{n}$$
• the observed proportion of data \( \leq X_{(i)} \), which is \( i/n \) (adjust to \( (i - 0.5)/n \) since it’s inconvenient to hit 0 or 1)

• If \( F(x) \) is the true (unknown) CDF of the data then \( F(x) = P(X \leq x) \) for any \( x \), so taking \( x = X_{(i)} \), \( F(X_{(i)}) = P(X \leq X_{(i)}) \), which is estimated by \( (i - 0.5)/n \)

• Thus, we should have \( F(X_{(i)}) \approx (i - 0.5)/n \), for all \( i = 1, 2, \ldots, n \)
• **P-P Plot**: If the fitted distribution (with CDF \( \hat{F} \)) is correct, i.e. close to the true unknown \( F \), we should have

\[
\hat{F}(X_{(i)}) \approx (i - 0.5) / n, \text{ for all } i = 1, 2, ..., n
\]

• so plotting the pairs \((i - 0.5) / n, \hat{F}(X_{(i)})\), for all \( i = 1, 2, ..., n \) should result in an approximately straight line from \((0, 0)\) to \((1, 1)\) if \( \hat{F} \) is correct

• Valid for both continuous and discrete data

• Sensitive to misfits in the center of the range of the distribution
• **Q-Q Plot:** Taking $\hat{F}^{-1}$ across the above,

$$\hat{F}^{-1}((i - 0.5) / n, X_{(i)}) \quad \text{for all } i = 1, 2, \ldots, n$$

• so plotting the pairs $((i-0.5) / n, \hat{F}(X_{(i)}))$, for all $i = 1, 2, \ldots, n$ should result in an approximately straight line from $(X_{(1)}, X_{(1)})$ to $(X_{(n)}, X_{(n)})$ if $\hat{F}$ is correct Valid only for continuous data

• Depending on the form of the fitted distribution, there may not be a closed-form formula for $\hat{F}^{-1}$

• Sensitive to misfits in the tails of the distributions
FIGURE 6.39
Definitions of $Q-Q$ and $P-P$ plots.
Distribution functions

$\tilde{F}_n(x)$

$\tilde{F}(x)$

$P-P$ plot

$\tilde{F}_n$ vs $\tilde{F}$

$Q-Q$ plot

$x^M$ vs $x^S$
<table>
<thead>
<tr>
<th>Distribution</th>
<th>Approach</th>
<th>Inverse Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gamma</td>
<td>See Bhattacharjee (1970)</td>
<td>See Best and Roberts (1975)</td>
</tr>
<tr>
<td>Lognormal</td>
<td>Fit a normal distribution to $Y_i = \ln X_i$ for $i = 1, 2, \ldots, n$; see Sec. 6.2.2</td>
<td>Same as $\hat{F}$</td>
</tr>
<tr>
<td>Beta</td>
<td>See Bosten and Battiste (1974)</td>
<td>See Cran, Martin, and Thomas (1977)</td>
</tr>
<tr>
<td>Pearson type V</td>
<td>Fit a gamma distribution to $Y_i = 1/X_i$ for $i = 1, 2, \ldots, n$; see Sec. 6.2.2</td>
<td>Same as $\hat{F}$</td>
</tr>
<tr>
<td>Pearson type VI</td>
<td>Fit a beta distribution to $Y_i = X_i/(1 + X_i)$ for $i = 1, 2, \ldots, n$; see Sec. 6.2.2</td>
<td>Same as $\hat{F}$</td>
</tr>
<tr>
<td>Johnson $S_B$</td>
<td>See the normal distribution</td>
<td>Same as $\hat{F}$</td>
</tr>
<tr>
<td>Johnson $S_U$</td>
<td>See the normal distribution</td>
<td>Same as $\hat{F}$</td>
</tr>
</tbody>
</table>
P-P plot of interarrival-time data for fitted exponential distribution:

Q-Q plot of interarrival-time data for fitted exponential distribution:

P-P plot of demand-size data for fitted geometric distribution:
6.6.2 Goodness-of-Fit Tests

- A Goodness-of-Fit Tests is a statistical hypothesis test
- Formal statistical hypothesis tests for
  \[ H_0: \text{the observed data } X_1, X_2, \ldots, X_n \text{ are IID random variables with distribution function } \hat{F} \]
- Caution: Failure to reject \( H_0 \) does not constitute “proof” that the fit is good. Power of some goodness-of-fit tests is low, particularly for small sample size \( n \)
- Also, large \( n \) creates high power, so tests will nearly always reject \( H_0 \).
Chi-Square Tests

- Very old (Karl Pearson, 1900), and general (continuous or discrete data) Formalization of frequency comparisons
- 1. Divide range of data into $k$ intervals, not necessarily of equal width:
  - $[a_0, a_1), [a_1, a_2), \ldots, [a_{k-1}, a_k)$
  - $a_0$ could be $-\infty$ or $a_k$ could be $+\infty$
2. Compare actual amount of observed data in each interval with what the fitted distribution would predict

- Let $N_j = \text{the number of observed data points in the } j\text{th interval}$
- Let $p_j = \text{the expected proportion of the data in the } j\text{th interval if the fitted distribution were literally true}$

\[
p_j = \begin{cases} 
\int_{a_{j-1}}^{a_j} \hat{f}(x) \, dx \text{ for continuous} \\
\sum_{a_{j-1} \leq x \leq a_j} \hat{p}(x) \text{ for discrete}
\end{cases}
\]
• Thus, \( n p_j = \text{expected (under fitted distribution) number of points in the } j \text{th interval} \)

• If fitted distribution is correct, would expect that \( N_j \approx n p_j \)

• Test statistic:

\[
\chi^2 = \sum_{j=1}^{k} \frac{(N_j - np_j)^2}{np_j}
\]

• Under \( H_0 \): fitted distribution is correct, \( \chi^2 \) has (approximately) a chi-square distribution with \( k-1 \) distribution function

• Reject \( H_0 \) at level \( \alpha \) if \( \chi^2 > \text{upper critical value} \)
Chi-square density with \( k - 1 \) df

Shaded area = \( \alpha \)

Do not reject

\( \chi^2_{k-1, 1-\alpha} \)

Reject

**FIGURE 6.45**
The chi-square test when all parameters are known.
Advantages:

- Completely general
- Asymptotically valid (as $n \to \infty$) if MLEs were used

Drawback:

- Arbitrary choice of intervals (can affect test conclusion)
- Conventional advice
  - Want $n p_j \geq 5$ or so for all but a couple of $j$’s
  - Pick intervals such that the $p_j$’s are close to each other
Asymptotic distribution function of $\chi^2$ if $H_0$ is true

$F_{k-m-1}(x)$

$F_{k-1}(x)$

$\chi^2_{k-m-1, 1-\alpha}$

$\chi^2_{1-\alpha}$

$\chi^2_{k-1, 1-\alpha}$

FIGURE 6.46

The chi-square test when $m$ parameters are estimated by their MLEs.
### Table 6.12
A chi-square goodness-of-fit test for the interarrival-time data

<table>
<thead>
<tr>
<th>Interval</th>
<th>( N_j )</th>
<th>( np_j )</th>
<th>( \frac{(N_j - np_j)^2}{np_j} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>([0, 0.020])</td>
<td>8</td>
<td>10.950</td>
<td>0.795</td>
</tr>
<tr>
<td>([0.020, 0.042])</td>
<td>11</td>
<td>10.950</td>
<td>0.000</td>
</tr>
<tr>
<td>([0.042, 0.065])</td>
<td>14</td>
<td>10.950</td>
<td>0.850</td>
</tr>
<tr>
<td>([0.065, 0.089])</td>
<td>14</td>
<td>10.950</td>
<td>0.850</td>
</tr>
<tr>
<td>([0.089, 0.115])</td>
<td>16</td>
<td>10.950</td>
<td>2.329</td>
</tr>
<tr>
<td>([0.115, 0.142])</td>
<td>10</td>
<td>10.950</td>
<td>0.082</td>
</tr>
<tr>
<td>([0.142, 0.172])</td>
<td>7</td>
<td>10.950</td>
<td>1.425</td>
</tr>
<tr>
<td>([0.172, 0.204])</td>
<td>5</td>
<td>10.950</td>
<td>3.233</td>
</tr>
<tr>
<td>([0.204, 0.239])</td>
<td>13</td>
<td>10.950</td>
<td>0.384</td>
</tr>
<tr>
<td>([0.239, 0.277])</td>
<td>12</td>
<td>10.950</td>
<td>0.101</td>
</tr>
<tr>
<td>([0.277, 0.319])</td>
<td>7</td>
<td>10.950</td>
<td>1.425</td>
</tr>
<tr>
<td>([0.319, 0.366])</td>
<td>7</td>
<td>10.950</td>
<td>1.425</td>
</tr>
<tr>
<td>([0.366, 0.519])</td>
<td>12</td>
<td>10.950</td>
<td>0.101</td>
</tr>
<tr>
<td>([0.419, 0.480])</td>
<td>10</td>
<td>10.950</td>
<td>0.082</td>
</tr>
<tr>
<td>([0.480, 0.553])</td>
<td>20</td>
<td>10.950</td>
<td>7.480</td>
</tr>
<tr>
<td>([0.553, 0.642])</td>
<td>9</td>
<td>10.950</td>
<td>0.347</td>
</tr>
<tr>
<td>([0.642, 0.757])</td>
<td>11</td>
<td>10.950</td>
<td>0.000</td>
</tr>
<tr>
<td>([0.757, 0.919])</td>
<td>9</td>
<td>10.950</td>
<td>0.347</td>
</tr>
<tr>
<td>([0.919, 1.195])</td>
<td>14</td>
<td>10.950</td>
<td>0.850</td>
</tr>
<tr>
<td>([1.195, \infty])</td>
<td>10</td>
<td>10.950</td>
<td>0.082</td>
</tr>
</tbody>
</table>

\( \chi^2 = 22.89 \)

### Table 6.13
A chi-square goodness-of-fit test for the demand-size data

<table>
<thead>
<tr>
<th>Interval</th>
<th>( N_j )</th>
<th>( np_j )</th>
<th>( \frac{(N_j - np_j)^2}{np_j} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>([0])</td>
<td>59</td>
<td>53.960</td>
<td>0.471</td>
</tr>
<tr>
<td>([1, 2])</td>
<td>50</td>
<td>58.382</td>
<td>1.203</td>
</tr>
<tr>
<td>([3, 4, \ldots])</td>
<td>47</td>
<td>43.658</td>
<td>0.256</td>
</tr>
</tbody>
</table>

\( \chi^2 = 1.930 \)
Chi-square test for exponential distribution fitted to interarrival-time data:

- Chose \( k = 20 \) intervals so that \( p_j = 1/20 = 0.05 \) for each interval
- Thus, \( np_j = (219) (0.05) = 10.95 \) for each interval
- Counted observed frequencies \( N_j \), computed test statistic \( \chi^2 = 22.188 \)
- Use d.f. = \( k-1 = 19 \); upper 0.10 critical level is \( \chi^2_{19,0.9} = 27.204 \)
- Since test statistic does not exceed the critical level, do not reject \( H_0 \)
Chi-square test for geometric distribution fitted to demand-size data:

- Since data are discrete, cannot choose intervals so that the $p_j$'s are exactly equal to each other
- Chose $k = 3$ intervals (classes) \{0\}, \{1, 2\}, and \{3, 4, ...\}
- Got $np_1 = 53.960$, $np_2 = 58.382$, and $np_3 = 43.658$
- Counted observed frequencies $N_j$, computed test statistic $\chi^2 = 1.930$
- Use d.f. = $k–1= 2$; upper 0.10 critical level is $\chi^2_{2,0.9} = 4.605$
- Since test statistic does not exceed the critical level, do not reject $H_0$