Review of Statistics: Bernoulli Distribution

Dr. Patrick Toche

Textbook:

James H. Stock and Mark W. Watson, *Introduction to Econometrics*, 4th Edition, Pearson.

Other references:

Joshua D. Angrist and Jörn-Steffen Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*, 1st Edition, Princeton University Press.
Jeffrey M. Wooldridge, *Introductory Econometrics: A Modern Approach*, 7th Edition, Cengage Learning.

The textbook comes with online resources and study guides. Other references will be given from time to time.

Problems and Applications: Bernoulli Distribution

Let Y be a Bernoulli random variable with success probability $\Pr(Y = 1) = p$, and let $Y_1, \ldots, Y_n$ be i.i.d. draws from this distribution. Let $\hat{p}$ be the fraction of successes (1s) in this sample.

1. Show that $\hat{p} = \overline{Y}$.

   $\hat{p}$ is the fraction of successes in this sample, so

   $$\hat{p} = \frac{1}{n} \cdot (Y_1 + \ldots + Y_n)$$

   The mean $\overline{Y}$ is the expected value of Y,

   $$\overline{Y} = \mathrm{E}[Y] = \frac{1}{n} \cdot (\mathrm{E}[Y_1] + \ldots \mathrm{E}[Y_n])$$

   And since $\hat{p}$ is a fixed sample value,

   $$\hat{p} = \mathrm{E}[\hat{p}] = \frac{1}{n} \cdot (\mathrm{E}[Y_1] + \ldots \mathrm{E}[Y_n]) = \overline{Y}$$

Problems and Applications: Bernoulli Distribution

Let Y be a Bernoulli random variable with success probability $\Pr(Y = 1) = p$, and let $Y_1, \ldots, Y_n$ be i.i.d. draws from this distribution. Let $\hat{p}$ be the fraction of successes (1s) in this sample.

1. Show that $\hat{p} = \overline{Y}$.

   $\hat{p}$ is the fraction of successes in this sample, so

   $$\hat{p} = \frac{1}{n} \cdot (Y_1 + \ldots + Y_n)$$

   The mean $\overline{Y}$ is the expected value of Y,

   $$\overline{Y} = \mathrm{E}[Y] = \frac{1}{n} \cdot (\mathrm{E}[Y_1] + \ldots \mathrm{E}[Y_n])$$

   And since $\hat{p}$ is a fixed sample value,

   $$\hat{p} = \mathrm{E}[\hat{p}] = \frac{1}{n} \cdot (\mathrm{E}[Y_1] + \ldots \mathrm{E}[Y_n]) = \overline{Y}$$

Problems and Applications: Bernoulli Distribution

Let Y be a Bernoulli random variable with success probability $\Pr(Y = 1) = p$, and let $Y_1, \ldots, Y_n$ be i.i.d. draws from this distribution. Let $\hat{p}$ be the fraction of successes (1s) in this sample.

2. Show that $\hat{p}$ is an unbiased estimator of p.

   The random variable Y_i takes value 1 with probability p and 0 with probability $1 - p$,

$$\mathrm{E}[Y_i] = p \cdot 1 + (1 - p) \cdot 0 = p$$

   By definition of the fraction of successes $\hat{p}$,

$$\begin{aligned} \hat{p} &= \frac{1}{n} \cdot (Y_1 + \ldots + Y_n) \\ \implies \mathrm{E}[\hat{p}] &= \frac{1}{n} \cdot (\mathrm{E}[Y_1] + \ldots + \mathrm{E}[Y_n]) \\ &= \frac{1}{n} \cdot (p + \ldots + p) \\ &= \frac{1}{n} \cdot n \cdot p = p \implies \mathrm{E}[\hat{p} - p] = 0 \end{aligned}$$

Problems and Applications: Bernoulli Distribution

Let Y be a Bernoulli random variable with success probability $\Pr(Y = 1) = p$, and let $Y_1, \ldots, Y_n$ be i.i.d. draws from this distribution. Let $\hat{p}$ be the fraction of successes (1s) in this sample.

2. Show that $\hat{p}$ is an unbiased estimator of p.

   The random variable Y_i takes value 1 with probability p and 0 with probability $1 - p$,

$$\mathrm{E}[Y_i] = p \cdot 1 + (1 - p) \cdot 0 = p$$

   By definition of the fraction of successes $\hat{p}$,

$$\begin{aligned} \hat{p} &= \frac{1}{n} \cdot (Y_1 + \ldots + Y_n) \\ \implies \mathrm{E}[\hat{p}] &= \frac{1}{n} \cdot (\mathrm{E}[Y_1] + \ldots + \mathrm{E}[Y_n]) \\ &= \frac{1}{n} \cdot (p + \ldots + p) \\ &= \frac{1}{n} \cdot n \cdot p = p \implies \mathrm{E}[\hat{p} - p] = 0 \end{aligned}$$

Problems and Applications: Bernoulli Distribution

Let Y be a Bernoulli random variable with success probability $\Pr(Y = 1) = p$, and let $Y_1, \ldots, Y_n$ be i.i.d. draws from this distribution. Let $\hat{p}$ be the fraction of successes (1s) in this sample.

3. Show that $\mathrm{var}(\hat{p}) = p(1 - p)/n$.

$$
\begin{aligned}
\mathrm{var}(\hat{p}) = \mathrm{var}(\overline{Y}) = \mathrm{var}\left(\frac{1}{n}(Y_1 + \ldots + Y_n)\right) & \\
= \frac{1}{n^2} \cdot \mathrm{var}(Y_1 + \ldots + Y_n) & \qquad \text{because } n \text{ is constant} \\
= \frac{1}{n^2} \cdot (\mathrm{var}(Y_1) + \ldots + \mathrm{var}(Y_n)) & \qquad \text{because the } Y_i\text{s are i.i.d.} \\
= \frac{1}{n^2} \cdot (\mathrm{var}(Y) + \ldots + \mathrm{var}(Y)) & \qquad \text{because the } Y_i\text{s are i.i.d.} \\
= \frac{1}{n^2} \cdot n \cdot \mathrm{var}(Y) & \\
= \frac{1}{n} \cdot \mathrm{var}(Y) &
\end{aligned}
$$

Problems and Applications: Bernoulli Distribution

Let Y be a Bernoulli random variable with success probability $\Pr(Y = 1) = p$, and let $Y_1, \ldots, Y_n$ be i.i.d. draws from this distribution. Let $\hat{p}$ be the fraction of successes (1s) in this sample.

3. Show that $\mathrm{var}(\hat{p}) = p(1-p)/n$.

$$\begin{aligned}
\mathrm{var}(\hat{p}) = \mathrm{var}(\overline{Y}) &= \mathrm{var}\left(\frac{1}{n}(Y_1 + \ldots + Y_n)\right) \\
&= \frac{1}{n^2} \cdot \mathrm{var}(Y_1 + \ldots + Y_n) \qquad \text{because } n \text{ is constant} \\
&= \frac{1}{n^2} \cdot (\mathrm{var}(Y_1) + \ldots + \mathrm{var}(Y_n)) \qquad \text{because the } Y_i\text{s are i.i.d.} \\
&= \frac{1}{n^2} \cdot (\mathrm{var}(Y) + \ldots + \mathrm{var}(Y)) \qquad \text{because the } Y_i\text{s are i.i.d.} \\
&= \frac{1}{n^2} \cdot n \cdot \mathrm{var}(Y) \\
&= \frac{1}{n} \cdot \mathrm{var}(Y)
\end{aligned}$$

Problems and Applications: Bernoulli Distribution

Let Y be a Bernoulli random variable with success probability $\Pr(Y = 1) = p$, and let $Y_1, \ldots, Y_n$ be i.i.d. draws from this distribution. Let $\hat{p}$ be the fraction of successes (1s) in this sample.

4. Show that $\mathrm{var}(\hat{p}) = p(1-p)/n$.

   Now compute $\mathrm{var}(Y)$

$$
\begin{aligned}
\mathrm{E}[Y] &= p \cdot 1 \;\; + (1-p) \cdot 0 \;\; = p \\
\mathrm{E}[Y^2] &= p \cdot 1^2 + (1-p) \cdot 0^2 = p \\
\mathrm{var}(Y) &= \mathrm{E}[Y^2] - (\mathrm{E}[Y])^2 = p - p^2 = p(1-p) \\
\implies \mathrm{var}(\hat{p}) &= \mathrm{var}(\overline{Y}) = \frac{\mathrm{var}(Y)}{n} = \frac{p(1-p)}{n}
\end{aligned}
$$

Problems and Applications: Bernoulli Distribution

Let Y be a Bernoulli random variable with success probability $\Pr(Y = 1) = p$, and let $Y_1, \ldots, Y_n$ be i.i.d. draws from this distribution. Let $\hat{p}$ be the fraction of successes (1s) in this sample.

4. Show that $\operatorname{var}(\hat{p}) = p(1-p)/n$.

   Now compute $\operatorname{var}(Y)$

$$\begin{aligned} \mathrm{E}[Y] &= p \cdot 1 \;\; + (1-p) \cdot 0 \;\; = p \\ \mathrm{E}[Y^2] &= p \cdot 1^2 + (1-p) \cdot 0^2 = p \\ \operatorname{var}(Y) &= \mathrm{E}[Y^2] - (\mathrm{E}[Y])^2 = p - p^2 = p(1-p) \\ \implies \operatorname{var}(\hat{p}) &= \operatorname{var}(\overline{Y}) = \frac{\operatorname{var}(Y)}{n} = \frac{p(1-p)}{n} \end{aligned}$$