Parameter Estimation

Least Squares Best Fit Line

C. Durso

Minimum Square-Error Parameters

Example

Suppose you are given n pairs of numbers, $\{(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)\}$ and you want to model y as a linear function of x, $y = mx + b$. You can choose the parameters m and b to minimize the sum of the squares of the errors of approximating y in this way:

$$\Sigma_{i=1}^{n}\left(y_i - (mx_i + b)\right)^2.$$

Differentiate

$$\frac{\partial}{\partial m}\left(\Sigma_{i=1}^{n}\left(y_i - (mx_i + b)\right)^2\right) = -\Sigma_{i=1}^{n} 2\left(y_i - (mx_i + b)\right)x_i$$

$$\frac{\partial}{\partial b}\left(\Sigma_{i=1}^{n}\left(y_i - (mx_i + b)\right)^2\right) = -\Sigma_{i=1}^{n} 2\left(y_i - (mx_i + b)\right)$$

Solve for b in Terms of m

$$-\Sigma_{i=1}^{n} 2\left(y_i - (mx_i + b)\right) = 0$$

Solve for b in Terms of m

$$\Sigma_{i=1}^{n} 2\big(y_i - (mx_i + b)\big) = 0$$

$$\Sigma_{i=1}^{n} y_i - m\Sigma_{i=1}^{n} x_i - nb = 0$$

Solve for b in Terms of m

$$\Sigma_{i=1}^{n} 2\big(y_i - (mx_i + b)\big) = 0$$

$$\Sigma_{i=1}^{n} y_i - m\Sigma_{i=1}^{n} x_i - \mathrm{n}b = 0$$

$$b = \bar{y} - m\bar{x}$$

Solve for m

Plug in the value of b in terms of m found above.

$$-2\Sigma_{i=1}^{n}\big(y_i - (mx_i + b)\big)x_i = -2\Sigma_{i=1}^{n}\big(y_i - mx_i - (\bar{y} - m\bar{x})\big)x_i = 0$$

Solve for m

Plug in the value of b in terms of m found above.

$$-2\Sigma_{i=1}^{n}\big(y_i - (mx_i + b)\big)x_i = -2\Sigma_{i=1}^{n}\big(y_i - mx_i - (\bar{y} - m\bar{x})\big)x_i = 0$$

$$\Sigma_{i=1}^{n}\big(y_i x_i - m x_i^2 - \bar{y}x_i + m\bar{x}x_i\big) = 0$$

Solve for m

Plug in the value of b in terms of m found above.

$$-2\Sigma_{i=1}^{n}\big(y_i - (mx_i + b)\big)x_i = -2\Sigma_{i=1}^{n}\big(y_i - mx_i - (\bar{y} - m\bar{x})\big)x_i = 0$$

$$\Sigma_{i=1}^{n}\big(y_i x_i - mx_i^2 - \bar{y}x_i + m\bar{x}x_i\big) = 0$$

$$(\Sigma_{i=1}^{n} y_i x_i - n\bar{y}\bar{x}) - m\big(\Sigma_{i=1}^{n} x_i^2 - n\bar{x}^2\big) = 0$$

Solve for m

Plug in the value of b in terms of m found above.

$$-2\Sigma_{i=1}^{n}\big(y_i - (mx_i + b)\big)x_i = -2\Sigma_{i=1}^{n}\big(y_i - mx_i - (\bar{y} - m\bar{x})\big)x_i = 0$$

$$\Sigma_{i=1}^{n}\big(y_i x_i - m x_i^2 - \bar{y}x_i + m\bar{x}x_i\big) = 0$$

$$(\Sigma_{i=1}^{n} y_i x_i - n\bar{y}\bar{x}) - m\big(\Sigma_{i=1}^{n} x_i^2 - n\bar{x}^2\big) = 0$$

$$m = \frac{\frac{1}{n}\Sigma_{i=1}^{n} y_i x_i - \bar{y}\bar{x}}{\frac{1}{n}\Sigma_{i=1}^{n} x_i^2 - \bar{x}^2}$$

Minimum Square Error Conclusion

Theorem

Given n pairs of numbers, $\{(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)\}$ and the model, $y = mx + b$, the parameters m and b that minimize the sum of the squares of the errors, $\Sigma_{i=1}^{n}\left(y_i - (mx_i + b)\right)^2$, are

- $m = \dfrac{\frac{1}{n}\Sigma_{i=1}^{n} y_i x_i - \bar{y}\bar{x}}{\frac{1}{n}\Sigma_{i=1}^{n} x_i^2 - \bar{x}^2}$
- $b = \bar{y} - m\bar{x}$ as above.

Mean and Variance of Sample Mean

C. Durso

Sample Mean

Definition

If $y_1, y_2, \dots y_n$ are numerical data values, the sample mean of $y_1, y_2, \dots y_n$ is $\frac{\sum_{i=1}^{n} y_i}{n}$, commonly denoted $\bar{y}$.

Sample Variance

Definition

If $y_1, y_2, \ldots y_n$ are numerical data values, the sample variance of $y_1, y_2, \ldots y_n$ is $\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}$ where $\bar{y}$ is the sample mean.

There is an alternate form of the sample variance that is better suited to streaming calculations.

$$\frac{\Sigma_{i=1}^{n}(y_i - \bar{y})^2}{n-1} = \frac{\Sigma_{i=1}^{n} y_i^2 - n\bar{y}^2}{n-1}$$

Alternate Form of Sample Variance

Theorem

The sample variance of $y_1, y_2, \dots y_n$ is equal to $\frac{\sum_{i=1}^{n} y_i^2 - n\bar{y}^2}{n-1}$.

$$\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1} = \frac{\sum_{i=1}^{n} y_i^2}{n-1} - \frac{\sum_{i=1}^{n} 2y_i\bar{y}}{n-1} + \frac{\sum_{i=1}^{n} \bar{y}^2}{n-1}$$

$$= \frac{\sum_{i=1}^{n} y_i^2}{n-1} - \frac{2\bar{y}\sum_{i=1}^{n} y_i}{n-1} + \frac{n\bar{y}^2}{n-1}$$

$$= \frac{\sum_{i=1}^{n} y_i^2}{n-1} - \frac{2\bar{y}n\bar{y}}{n-1} + \frac{n\bar{y}^2}{n-1} = \frac{\sum_{i=1}^{n} y_i^2 - n\bar{y}^2}{n-1}$$

Population Mean and Variance

C. Durso

Population or Distribution Mean

Definition

For a discrete random variable $Y = (S \subseteq \mathbb{R}, M, P)$ where P is induced by a density f, the *mean* of Y equals $\sum_{y \in S} y f(y)$. It is denoted $\bar{Y}$.

Definition

For a continuous random variable $Y = (S \subseteq \mathbb{R}, M, P)$ where P is induced by a density f, the *mean* of Y equals $\int y f(y) dy$. It is denoted $\bar{Y}$.

Note that the mean may not be well-defined. The sum or integral may not converge.

Population or Distribution Variance

Definition

For a discrete random variable $Y = (S \subseteq \mathbb{R}, M, P)$ where P is induced by a density f, the *variance* of Y equals $\sum_{y \in S}(y - \bar{Y})^2 f(y)$.

Definition

For a continuous random variable $Y = (S \subseteq \mathbb{R}, M, P)$ where P is induced by a density f, the *variance* of Y equals $\int (y - \bar{Y})^2 f(y) dy$.

Note that the variance may not be well-defined. The sum or integral may not converge.

Alternate Form of Variance of a Distribution

Theorem

For a discrete random variable $Y = (S \subseteq \mathbb{R}, M, P)$ where P is induced by a density f, the *variance* of Y equals $\sum_{y \in S} y^2 f(y) - \bar{Y}^2$.

Theorem

For a continuous random variable $Y = (S \subseteq \mathbb{R}, M, P)$ where P is induced by a density f, the *variance* of Y equals $\int y^2 f(y) dy - \bar{Y}^2$.

Proof of Alternate Form

The proof is provided for the discrete case. The continuous case is similar.

$$\sum_{y \in S} (y_i - \bar{Y})^2 f(y) = \sum_{y \in S} y_i^2 f(y) - \sum_{y \in S} 2 y_i \bar{Y} f(y) + \sum_{y \in S} \bar{Y}^2 f(y)$$

$$= \sum_{y \in S} y_i^2 f(y) - 2\bar{Y} \sum_{y \in S} y_i f(y) + \bar{Y}^2$$

$$= \sum_{y \in S} y_i^2 f(y) - 2\bar{Y}^2 + \bar{Y}^2 = \sum_{y \in S} y_i^2 f(y) - \bar{Y}^2$$

Binomial Expected Value

Example

C. Durso

Expected Value of Binomial(n, p)

Formula

$$\sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = np$$

Why?

$$\sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k}$$

$$\sum_{k=1}^{n} k \frac{n!}{(n-k)!\,k!} p^k (1-p)^{n-k}$$

$$\sum_{k=1}^{n} n \frac{(n-1)!}{\big(n-1-(k-1)\big)!\,(k-1)!} p^k (1-p)^{n-k}$$

$$\sum_{k=1}^{n} np \frac{(n-1)!}{\big(n-1-(k-1)\big)!\,(k-1)!} p^{k-1} (1-p)^{(n-1-(k-1))}$$

$$np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-1-k} = np$$

There’s an Easier Way

Watch for identities for the expected value of sums that will make this computation much simpler.

Binomial Variance

Example

C. Durso

Variance of Binomial(n,p)

Formula

$$\sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k} - \left(\sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k}\right)^2 = np(1-p)$$

Derivation Plan

- Known:

$$\sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = np$$

- To do:

$$\sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k} = np(1-p) + n^2p^2$$

- Rewrite:

$$\sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k}$$

as

$$\sum_{k=0}^{n} [k(k-1)+k] \binom{n}{k} p^k (1-p)^{n-k}$$

Derivation Details

$$\sum_{k=0}^{n}[k(k-1)+k]\binom{n}{k}p^k(1-p)^{n-k} =$$

$$\sum_{k=0}^{n}k(k-1)\binom{n}{k}p^k(1-p)^{n-k}+np =$$

$$\sum_{k=2}^{n}n(n-1)\binom{n-2}{k-2}p^k(1-p)^{(n-2)-(k-2)}+np =$$

$$n(n-1)p^2\sum_{k=2}^{n}\binom{n-2}{k-2}p^{k-2}(1-p)^{(n-2)-(k-2)}+np =$$

$$n(n-1)p^2\sum_{k=0}^{n-2}\binom{n-2}{k}p^k(1-p)^{(n-2)-k}+np =$$

$$n(n-1)p^2+np = np(1-p)+n^2p^2$$

as required.

Normal(μ, σ^2) Expected Value

Example

C. Durso

Expected Value of Normal(μ, σ^2)

Formula

$$\int_{-\infty}^{\infty} \frac{x}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = \mu$$

For General μ

$$\int_{-\infty}^{\infty} \frac{x}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx$$

$$\int_{-\infty}^{\infty} \frac{u+\mu}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{u^2}{2\sigma^2}\right) du$$

$$\int_{-\infty}^{\infty} \frac{u}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{u^2}{2\sigma^2}\right) du + \int_{-\infty}^{\infty} \frac{\mu}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{u^2}{2\sigma^2}\right) du$$

$$0 + \mu$$

Normal(μ, σ^2) Variance

Example

C. Durso

Variance of Normal(μ, σ^2)

Formula

$$\int_{-\infty}^{\infty} \frac{x^2}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx - \left[\int_{-\infty}^{\infty} \frac{x}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx\right]^2 = \sigma^2$$

Variance of Normal$(0,1)$

- We know that the mean of Normal$(0,1)$ is 0, so we need to verify that
$$\int_{-\infty}^{\infty} \frac{x^2}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx = 1$$
- Use integration by parts with $u = x$ and $dv = x\exp\left(-\frac{x^2}{2}\right), v = -\exp\left(-\frac{x^2}{2}\right)$.

Variance of Normal$(0, \sigma^2)$

- Given $\int_{-\infty}^{\infty} \frac{x^2}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx = 1$, we can see that $\int_{-\infty}^{\infty} \frac{x^2}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx = \sigma^2$ using the change of variable $u = \frac{x}{\sigma}$.
- This will enable us to conclude that the variance of Normal$(0, \sigma^2)$ equals σ^2.

Variance of Normal(μ, σ^2)

- We know that the expected value of Normal(μ, σ^2) equals μ, so the last piece is verification that $\int_{-\infty}^{\infty} \frac{x^2}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = \sigma^2 + \mu^2$.
- This follows from the change of variable $u = x - \mu$.

Variance of Normal(μ, σ^2)

Theorem

The variance of the normal distribution with parameters μ and σ^2 equals σ^2.

That is, $\mathrm{Var}[normal(\mu, \sigma^2)] = \sigma^2$