Outline

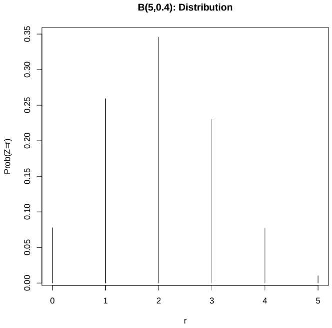
Binomial Distribution: Probability

- Suppose $Y \sim B(n, \pi)$ with probability density function

$$P(Y = y) = \frac{n!}{y!(n-y)!}\pi^y(1-\pi)^{n-y},$$

where $y = 0, 1, \ldots, n$.

- For example, $n = 5$ and $\pi = 0.4$. Plot $P(Y = y)$ versus y:



Binomial Distribution: Statistics

- Suppose there are $n = 5$ trials and the observed number of successes is $y = 2$.
- Q: How to estimate π?
- A method of moment (MOM) estimator is

Likelihood Function

- Consider the probability mass function evaluated at $y = 2$:

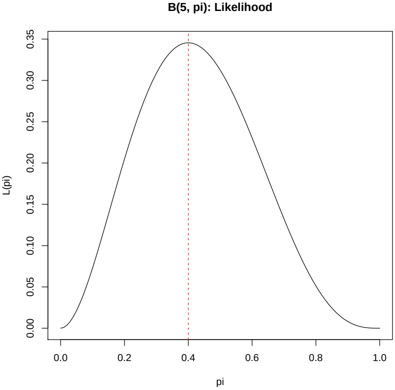
$$P(Y = 2) = \frac{5!}{2!3!}\pi^2(1 - \pi)^3.$$

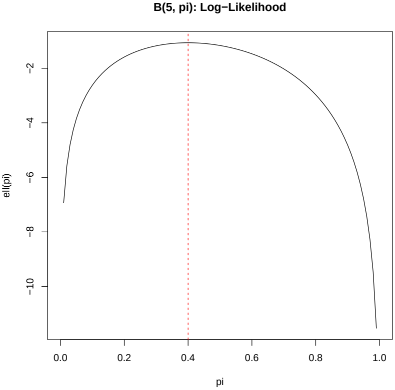
- Thus, we have

π	0.2	0.4	0.6	0.8
$P(Y = 2)$	0.2048	0.3456	0.2304	0.0512

Likelihood Function

- Plot $P(Y = 2)$ versus $\pi = 0.01, 0.02, \ldots, 0.98, 0.99$:





- Q: What value of π makes the given data most likely?

Likelihood Function

- That is, find the value of π that maximizes

$$\frac{5!}{2!3!}\pi^2(1-\pi)^3$$

- Given n and y, the function

$$\mathcal{L}(\pi) = \frac{n!}{y!(n-y)!}\pi^y(1-\pi)^{n-y}$$

  is the **likelihood function** of the unknown parameter π.
- Further, the **log-likelihood function** of π is:

$$\ell(\pi) = y\ln(\pi) + (n-y)\ln(1-\pi) + \ln\left\{\frac{n!}{y!(n-y)!}\right\}.$$

Maximum Likelihood (ML) Estimation

- To maximize the log-likelihood function, set the derivative to 0 and solve for π:

$$\frac{d\ell(\pi)}{d\pi} = \frac{y}{\pi} - \frac{n-y}{1-\pi} = 0.$$

- The **maximum likelihood estimate (MLE)** of π is:

- The maximum log-likelihood value is

$$\begin{aligned}
\ell(\hat{\pi}) &= y\ln(\hat{\pi}) + (n-y)\ln(1-\hat{\pi}) + \ln\left\{\frac{n!}{y!(n-y)!}\right\} \\
&= 2 \times \ln\left(\frac{2}{5}\right) + 3 \times \ln\left(\frac{3}{5}\right) + \ln\left(\frac{5!}{2!3!}\right) \\
&= -1.0625
\end{aligned}$$

Definition (MLE)

The MLE for a parameter θ is the statistics $\hat{\theta} = T(y)$ whose value for the given data y satisfies the condition

$$L(\hat{\theta}|y) = \sup_{\theta \in \Theta} L(\theta|y),$$

where $L(\theta|y)$ is the likelihood function for θ.

Properties:

- MLEs are invariant; i.e., $MLE(g(\theta)) = g(MLE(\theta)) = g(\hat{\theta})$.
- MLEs are asymptotically normal and asymptotically unbiased.

Gaussian Distribution: ML Estimation

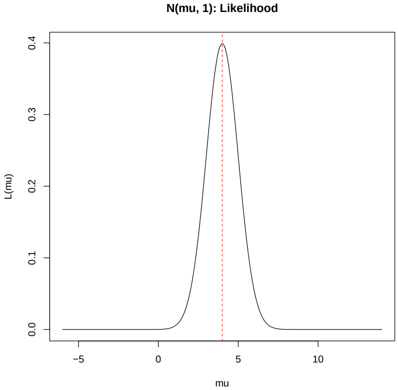
- Suppose $Y \sim N(\mu, 1)$ (i.e., $\sigma^2 = 1$ is known).
- Given the data $y = 4$, the maximum likelihood estimate (MLE) of μ is:
- The likelihood function of μ is:

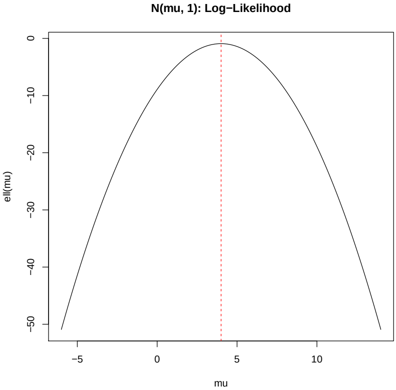
$$\mathcal{L}(\mu) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(4-\mu)^2\right\}.$$

- The log-likelihood function of μ is:

$$\ell(\mu) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}(4-\mu)^2.$$

Gaussian Distribution: ML Estimation





Point Estimation

A good estimate $\hat{\theta}$ should

- Be unbiased: $\mathbb{E}(\hat{\theta}) = \theta$
- Have small variance: small $\mathrm{Var}(\hat{\theta})$
- Be efficient: its mean squared error (MSE) is minimum among all competitors.

$$\mathrm{MSE}(\hat{\theta}) \equiv \mathbb{E}(\hat{\theta} - \theta)^2 = \mathrm{Bias}^2(\hat{\theta}) + \mathrm{Var}(\hat{\theta}),$$

  where $\mathrm{Bias}(\theta) = \mathbb{E}(\hat{\theta}) - \theta$.

- Be consistent:

$$\hat{\theta} = \hat{\theta}(n) \to \theta \quad \text{in probability, as the sample size} \quad n \to \infty.$$

Comparison

Method of Moment:

- Pros: easy to compute, consistent
- Cons: not necessarily the most efficient estimate; sometimes outside the valid range; may not be unique.

Maximum likelihood estimator:

- Pros: **asymptotically** unbiased, consistent, normally distributed, and efficient
- Cons: can be highly biased for small samples; sometimes, MLE has no closed-form.

Outline

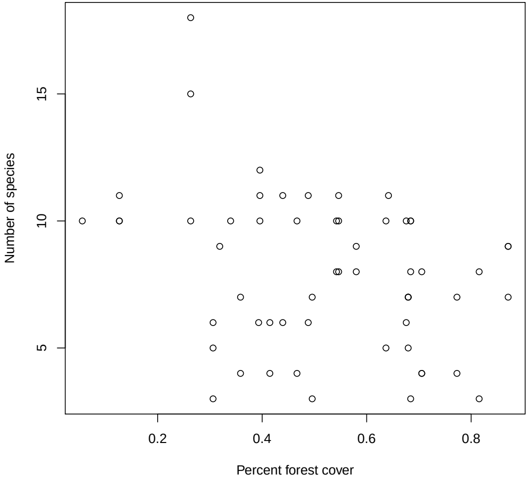
Example: Wetland Species Richness

- A study was performed on insect species richness in 58 wetlands in Ontario, Canada.
- The goal of the study was to determine the relationship between forest density around the wetland and insect species richness.
- The investigators sample insects in each wetland and then recorded the number of species present in each sample.
- The percent forest cover within a 1500-meter buffer around the wetland was also recorded, among other wetland characteristics.

Example: Wetland Species Richness

wetland	y	x	wetland	y	x
1	10	0.056	30	5	0.637
2	8	0.546	31	6	0.488
3	10	0.637	32	9	0.580
4	8	0.815	33	4	0.705
5	10	0.676	34	11	0.439
6	9	0.871	35	8	0.705
7	4	0.467	36	5	0.680
8	3	0.684	37	10	0.396
9	3	0.496	38	10	0.467
10	4	0.415	39	5	0.306
11	7	0.680	40	10	0.684
12	7	0.773	41	6	0.415
13	9	0.319	42	10	0.684
14	10	0.127	43	10	0.340
15	3	0.306	44	7	0.871
16	6	0.676	45	9	0.871
17	8	0.684	46	7	0.680
18	10	0.546	47	18	0.263
19	10	0.542	48	12	0.396
20	15	0.263	49	6	0.306
21	11	0.488	50	4	0.359
22	7	0.359	51	6	0.439
23	7	0.680	52	8	0.542
24	6	0.393	53	4	0.705
25	4	0.773	54	11	0.127
26	3	0.815	55	7	0.496
27	11	0.642	56	10	0.263
28	8	0.580	57	10	0.127
29	11	0.396	58	11	0.546

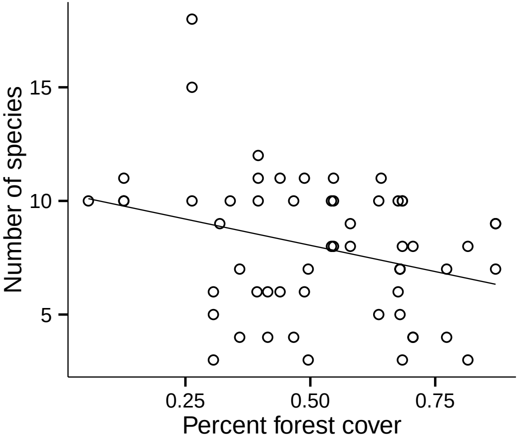
Example: Wetland Species Richness



Specific Goals

- To describe the relationship between the percent forest cover (x) and the number of species (y).
- To estimate or predict the number of species for a given percent forest cover.

Example: Wetland Species Richness



Q: How to account for uncertainty in the fitted line and variation?

Outline

Modeling Idea

- Model y by a random variable Y.
- Regard x as fixed, or condition on x (x could be modeled by a random variable X.)
- Consider the model of Y conditional on $X = x$:

$$E(Y|X = x) = \beta_0 + \beta_1 x.$$

- β_0, β_1 are fixed unknown parameters (i.e., the intercept and slope) characterizing the relationship between X and Y.

Simple Linear Regression Model

The formal simple linear regression (SLR) model for the data (x_i, y_i) is:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

for $i = 1, 2, \ldots, n$, where

- Y_i is the ith **response variable.**
- X_i is the ith **explanatory variable** (also called predictors, covariates).
- ε_i is the ith **random error** term.
  - The random errors follow a normal distribution with mean zero and variance σ^2 and are independent of each other.
  - That is, $\varepsilon_i \overset{\text{iid}}{\sim} N(0, \sigma^2)$.

Features of Simple Linear Regression Model

Under the SLR model for the data (x_i, y_i):

- Simple
- Linear
- Regression
- Randomness
- Independence
- The model parameters are:

Model Assumptions

- A straight line relationship between the response variable Y and the explanatory variable X:

$$E(Y_i|X_i) = \beta_0 + \beta_1 x_i.$$

- Equal variance:

$$Var(Y_i|X_i) = \sigma^2.$$

- Independence (conditional on X_i, X_i'):

$$Cov(Y_i, Y_{i'}) = 0 \quad \text{for} \quad i \neq i'.$$

- Normal distribution:

$$Y_i|X_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2).$$

Equivalent Model Assumptions

Equivalently, the assumptions are

- A straight line relationship between the response variable Y and the explanatory variable X:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad \text{where} \quad E(\varepsilon_i) = 0$$

- Equal variance:

$$Var(\varepsilon_i) = \sigma^2.$$

- Independence:

$$Cov(\varepsilon_i, \varepsilon_{i'}) = 0 \quad \text{for} \quad i \neq i'.$$

- Normal distribution:

$$\varepsilon_i \sim N(0, \sigma^2).$$

Model Parameters

- The model parameters are β_0, β_1, and σ^2 (population parameters).
- β_0 and β_1: **regression coefficients.**
- β_0: **intercept.**
  When the model scope includes $x = 0$, β_0 can be interpreted as the mean of Y at $x = 0$.
- β_1: **slope.**
  Interpreted as the change in the mean of Y per unit increase in x.
- σ^2: **error variance**, sometimes written as σ^2_ε or $\sigma^2_{Y|x}$.

Outline

Estimation of Model Parameters

- Our goal is to estimate these model parameters by estimators $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\sigma}^2$, based on data.
- Two methods:
  - Least squares (LS).
  - Maximum likelihood (ML).
- Additional notation:
  - Let $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ denote the *i*th fitted value.
  - Let $e_i = Y_i - \hat{Y}_i$ denote the *i*th residual.

Estimation of β_0 and β_1

- Both LS and ML give the same estimator for β_0 and β_1:

$$
\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \\
\hat{\beta}_0 &= \frac{1}{n}\left(\sum_{i=1}^{n} Y_i - \hat{\beta}_1 \sum_{i=1}^{n} X_i\right) = \bar{Y} - \hat{\beta}_1 \bar{X}.
\end{aligned}
$$

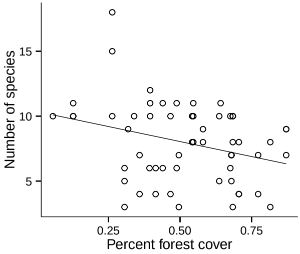
- Note: $\hat{\beta}_0$ and $\hat{\beta}_1$ are denoted as b_0 and b_1 in some texts.
- We now give **two methods** for these estimations.

Least Squares (LS) Estimation

- Consider the criterion:

$$Q = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2.$$

- The LS estimators of β_0 and β_1 are those values, $\hat{\beta}_0$ and $\hat{\beta}_1$, that minimize Q, for the given observed data $(X_1, Y_1), \ldots, (X_n, Y_n)$.
- Graphical interpretation?



LS Derivation

- Differentiate Q with respect to β_0 and β_1:

$$
\begin{aligned}
\text{(a)} \quad &: \quad \frac{\partial Q}{\partial \beta_0} = -2\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i) \\
\text{(b)} \quad &: \quad \frac{\partial Q}{\partial \beta_1} = -2\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)X_i
\end{aligned}
$$

- Set (a) and (b) equal to 0 and let the solutions to these two equations be $\hat{\beta}_0$ and $\hat{\beta}_1$.
- Let $\boldsymbol{\beta} = (\beta_0, \beta_1)'$.
- Since $\frac{\partial^2 Q}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}$ is positive definite, $\hat{\beta}_0$ and $\hat{\beta}_1$ minimize Q.

Gaussian Distribution: ML Estimation

- Suppose $Y_1, Y_2, \ldots, Y_n \sim$ iid $N(\mu, \sigma^2)$.
- Given the data $y_1, y_2, \ldots, y_n$, the likelihood function of μ, σ^2 is

$$\mathcal{L}(\mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\right\}.$$

- The log-likelihood function of μ, σ^2 is

$$\ell(\mu, \sigma^2) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2.$$

- The maximum likelihood estimate (MLE) for μ, σ^2 are:

General Distribution: ML Estimation

- In a general setting, let $Y_1, \ldots, Y_n$ be iid with probability density function $f(y; \boldsymbol{\theta})$.
- With $\boldsymbol{y} = (y_1, \ldots, y_n)'$, the likelihood function for $\boldsymbol{\theta}$ is

$$\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y}) = \prod_{i=1}^{n} f(y_i; \boldsymbol{\theta}).$$

- Find the value of $\boldsymbol{\theta}$ that maximizes $\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y})$.
- Equivalently, find the value of $\boldsymbol{\theta}$ that maximizes the log-likelihood

$$\ell(\boldsymbol{\theta}; \boldsymbol{y}) = \log \mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y}) = \log \prod_{i=1}^{n} f(y_i; \boldsymbol{\theta}) = \sum_{i=1}^{n} \log f(y_i; \boldsymbol{\theta}).$$

- Intuition:

ML Derivation

- Let $\boldsymbol{\theta} = (\beta_0, \beta_1, \sigma^2)'$.
- We have $Y_i \sim \mathrm{ind} N(\beta_0 + \beta_1 X_i, \sigma^2)$.
- Thus,

$$f_i(y_i; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}\left\{y_i - (\beta_0 + \beta_1 x_i)\right\}^2\right].$$

- The likelihood function is

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y}) &= \prod_{i=1}^{n} f_i(y_i; \boldsymbol{\theta}) \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}\left\{y_i - (\beta_0 + \beta_1 x_i)\right\}^2\right]\end{aligned}$$

ML Derivation

- The log-likelihood function is

$$
\begin{aligned}
\ell(\boldsymbol{\theta};\boldsymbol{y}) &= \sum_{i=1}^{n} \log f_i(y_i;\boldsymbol{\theta}) \\
&= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\{y_i - (\beta_0 + \beta_1 x_i)\}^2
\end{aligned}
$$

- Set the first-order partial derivatives equal to 0:

$$
\begin{aligned}
0 &= \frac{\partial\ell(\boldsymbol{\theta};\boldsymbol{y})}{\partial\beta_0} = \frac{2}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i) \\
0 &= \frac{\partial\ell(\boldsymbol{\theta};\boldsymbol{y})}{\partial\beta_1} = \frac{2}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)x_i \\
0 &= \frac{\partial\ell(\boldsymbol{\theta};\boldsymbol{y})}{\partial\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2.
\end{aligned}
$$

ML Derivation

Solve for the parameters and obtain the ML estimates:

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \\
\hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1 \bar{X} \\
\tilde{\sigma}^2 &= \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n}
\end{aligned}$$

Properties of Fitted Regression Line

For the fitted values $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ and residuals $e_i = Y_i - \hat{Y}_i$, we have:

- The regression line always goes through $(\bar{X}, \bar{Y})$.
- $\sum_{i=1}^n e_i^2$ is a minimum.
- $\sum_{i=1}^n e_i = 0$.
- $\sum_{i=1}^n X_i e_i = 0$.
- $\sum_{i=1}^n Y_i = \sum_{i=1}^n \hat{Y}_i$.
- $\sum_{i=1}^n \hat{Y}_i e_i = 0$.

Estimation of σ^2

- Define an **error sum of squares (SSE)** (or, **residual sum of squares**):

$$\mathrm{SSE} = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n} e_i^2 .$$

- Under simple linear regression, an unbiased estimate of σ^2 is an **error mean square (MSE)** (or, **residual mean square**):

$$\hat{\sigma}^2 = \mathrm{MSE} = \frac{\mathrm{SSE}}{n-2} = \frac{\sum_{i=1}^{n} e_i^2}{n-2} .$$

- The ML estimate of σ^2 is:

$$\tilde{\sigma}^2 = \frac{\mathrm{SSE}}{n} = \frac{\sum_{i=1}^{n} e_i^2}{n} .$$

Example: Wetland Species Richness

- In the wetland species richness example, we have

$$\mathrm{SSE} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} e_i^2 = 479.04$$

- Under LS, we have

$$\hat{\sigma}^2 = \mathrm{MSE} = \frac{\mathrm{SSE}}{n-2} = \frac{479.04}{56} = 8.554$$

- Under ML, we have

$$\tilde{\sigma}^2 = \frac{\mathrm{SSE}}{n} = \frac{479.04}{58} = 8.259.$$

- Which estimator is better?