Systems Genetics 02 - Primer in statistical modeling - Exercises

Julien Gagneur

26 October 2021

Package

BiocStyle 2.20.2

Contents

1 Maximum likelihood: Tossing coins

Consider n independent random tosses of a coin. We denote $x_i \in \{0;1\}$ the outcome of the i-th toss (1 for head and 0 for tail) and p the probablity to get a head.

Question 1: What is the maximum likelihood estimate of p? Prove it.

The R snippet `sample( c(0,1), size=n, prob=c(1-p,p), replace=TRUE)` draws `n` realizations of single tosses with probability `p` to return 1.

Question 2: Building on the code snippet above, implement a simulator and a max-likelihood estimator in R. Using simulations with various sample sizes n and probabilities p, investigate empirically the bias (is it on average on target?) and robustness (how far is it from the true p) of the ML estimator.

2 Gaussian linear systems

2.1 Marginalization

Assume:

$$p(x) = \mathcal{N}(x|a, \sigma_1^2) \tag{1}$$
$$p(y|x) = \mathcal{N}(y|x + b, \sigma_2^2) \tag{2}$$

Question 3: In R, simulate 10^3 random draws of $p(y)$ according to this model for various values $a, b, \sigma_1^2, \sigma_2^2$ of your choice. Check with normal quantile plots that $p(y)$ is normal and that its mean and variance depend on $a, b, \sigma_1^2, \sigma_2^2$ as expected by the relevant result(s) from the course.

Hint: The function `rnorm()` performs random draws according to the normal distribution. The calls `qqnorm(y)` and `qqline(y)` draw Q-Q plots against the normal distribution and the line of expected quantiles.

2.2 Conditioning

Assume x and y are 1-dimensional variables such that:

$$p(x) = \mathcal{N}(x|\mu_x, \sigma_x^2) \tag{3}$$
$$p(y|x) = \mathcal{N}(y|a\,x + b, \sigma_y^2) \tag{4}$$

for $a, b \in \mathbb{R}$.

Question 4: Show that $\mu_{x|y} := \mathrm{E}[x|y]$ is a convex combination of $(y-b)/a$ and μ_x:

$$\mu_{x|y} = w\frac{y-b}{a} + (1-w)\,\mu_x \tag{5}$$

What values of the parameter a, σ_x^2, and σ_y^2 can lead to the extreme cases $w \to 0$ and $w \to 1$? Interpret.