INSE6300 — Jia Yuan Yu

5: Parametric Estimation

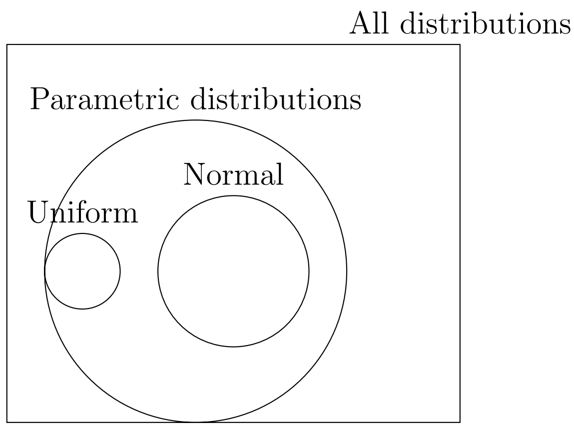
Concordia — October 21, 2017

Remark 1 (DKW-like inequalities). For other example of inequalities like the DKW Inequality, see "Understanding Machine Learning" Chapter 31, and Appendix B.4.

Previously, we have considered the problem of estimating the distribution of a random variable using the DFW Inequality, and the mean and variance of a random variable in the context of control charts.

Suppose that we know that the unknown distribution F_θ of the data $X_1, X_2, \dots$ belongs to a set $\{F_\gamma : \gamma \in \Omega\}$. How can we estimate the parameter θ or a function $g(\theta)$ of this parameter. Can we do this more efficiently than applying the DFW Inequality?

There are many methods for estimation: Bayesian, Max-likelihood, unbiased, etc. We overview some of these.



1 Unbiased estimation

A function δ is unbiased estimator for $g(\theta)$ if

$$\mathbb{E}_\theta \delta(X) = g(\theta), \quad \text{for all } \theta \in \Omega.$$

The bias is the estimation error.

1.1 Support of uniform distribution

Suppose that we want to estimate the parameter θ of the support $[0, \theta]$ of a uniform random variable. An unbiased estimation satisfies:

$$\mathbb{E}_\theta \delta(X) = \int_0^\theta \delta(x) \frac{1}{\theta} \mathrm{d}x = g(\theta), \quad \theta \geq 0,$$

or

$$\int_0^\theta \delta(x)\mathrm{d}x = \theta g(\theta), \quad \theta \geq 0.$$

Suppose that g is differentiable: by the FTC, we have

$$\delta(x) = \frac{\mathrm{d}}{\mathrm{d}x} x g(x) = g(x) + x g'(x).$$

Remark 2. If g is the identity, then $\delta(x) = 2x$ is an unbiased estimator for θ.

1.2 Binomial distribution

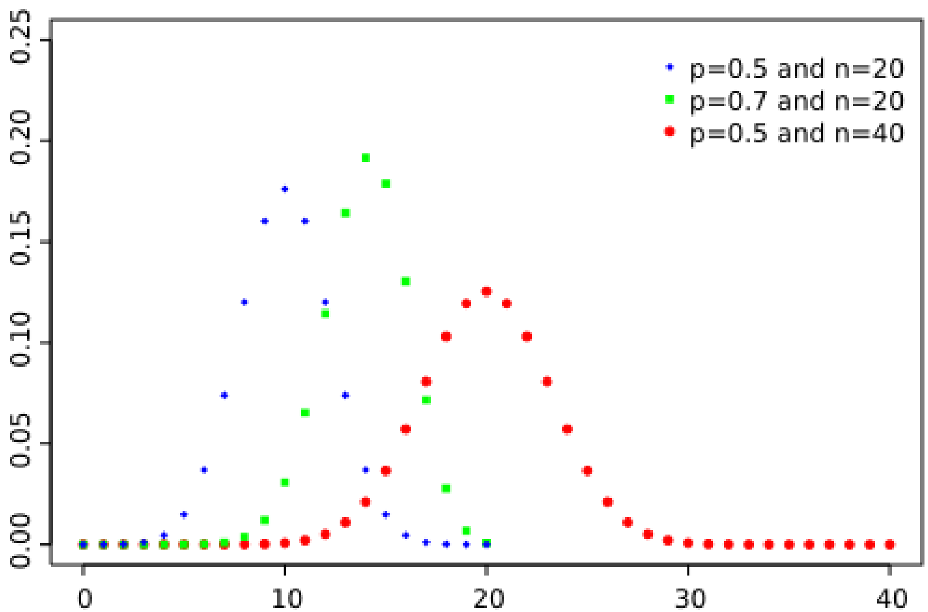


Figure 1: From Wikipedia.

The binomial distribution with parameters n (natural number) and $\theta \in [0, 1]$ is the PMF of a random variable counting the number of successes in n trials, probability of success θ. This distribution has mean $n\theta$ and variance $n\theta(1-\theta)$.

Suppose that we want to an unbiased estimator for $g(\theta) = \theta(1-\theta)$. We need to satisfy:

$$\mathbb{E}_\theta \delta(X) = \sum_{k=0}^{n} \binom{n}{k} \theta^k (1-\theta)^{n-k} \delta(k) = \theta(1-\theta), \quad \theta \geq 0. \tag{1}$$

Introduce $r = \theta/(1-\theta)$, we get

$$\theta^k (1-\theta)^{n-k} = r^k (1-\theta)^k \frac{\theta^{n-k}}{r^{n-k}} = \frac{r^k (1-\theta)^k \theta^n r^k}{\theta^k r^n} = \frac{r^k \theta^n}{r^n}.$$

Equation 1 then becomes:

$$\sum_{k=0}^{n} \binom{n}{k} r^k \delta(k) = \theta(1-\theta)\frac{r^n}{\theta^n} = r(1+r)^{n-2} = r\sum_{k=0}^{n-2}\binom{n-2}{k} r^k$$
$$= \sum_{k=1}^{n-2}\binom{n-2}{k-1} r^k,$$

where we used the Binomial theorem. Hence, an unbiased estimator for $\theta(1-\theta)$ is

$$\delta(k) = \frac{k(n-k)}{n(n-1)}.$$

Q: What about the unbiased estimator for θ? Can it be obtained by solving $\delta(k) = \theta(1-\theta)$.

1.3 Normal distribution

Let $X_1, \ldots, X_n$ denote measurements of the quality of n items. These are i.i.d. from a normal distribution with unknown mean μ and unknown variance σ^2. We are given a probability p, and we want to estimate the threshold u such that we can guarantee:

$$\mathbb{P}(X_{n+1} \leq u) = p.$$

Recall that

$$\mathbb{P}(X_{n+1} \leq u) = \Phi\left(\frac{u-\mu}{\sigma}\right),$$

so that

$$u = \mu + \sigma\Phi^{-1}(p).$$

Recall that the unbiased estimator for μ is the sample mean $\bar{X}$. The unbiased estimator for the variance is the sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$. However, the unbiased estimator for σ is not S! The unbiased estimator for σ is[1]

$$\left(\frac{n-1}{2}\right)^{1/2}\frac{\Gamma((n-1)/2)}{\Gamma(n/2)}S.$$

Therefore, the unbiased estimator for u is

$$\bar{X}_n + \left(\frac{n-1}{2}\right)^{1/2}\frac{\Gamma((n-1)/2)}{\Gamma(n/2)}S\Phi^{-1}(p),$$

where Γ is the Gamma function, which appears in various probability distributions (e.g., gamma and χ^2).

[1]See Robert W. Keener's "Theoretical Statistics: Topics for a Core Course," Chapter 4.4.

2 Maximum likelihood estimation

Let $X_1, X_2, \ldots, X_n$ be i.i.d. with probability density function f_θ for an unknown $\theta \in \Omega$. For a given $\omega \in \Omega$, the likelihood function is the product of the density f_ω evaluated at the data points:

$$\prod_{i=1}^{n} f_\omega(X_i).$$

The maximum likelihood estimator is:

$$\hat{\theta}_n \in \arg\max_{\omega \in \Omega} \prod_{i=1}^{n} f_\omega(X_i).$$

This is in general a random variable. It can be computed once the data is observed.

Homework: How does $\hat{\theta}_n$ compare with the unknown θ? Try on simulated random variables.

Example 2.1 (Binomial distribution)**.** Consider a binomial random variable X with unknown parameter θ and known parameter n. The likelihood function is

$$\binom{n}{X} \theta^X (1-\theta)^{n-X}.$$

We can plot the above likelihood function as a function of p and solve for the maximum likelihood estimate (cf. Figure 1).

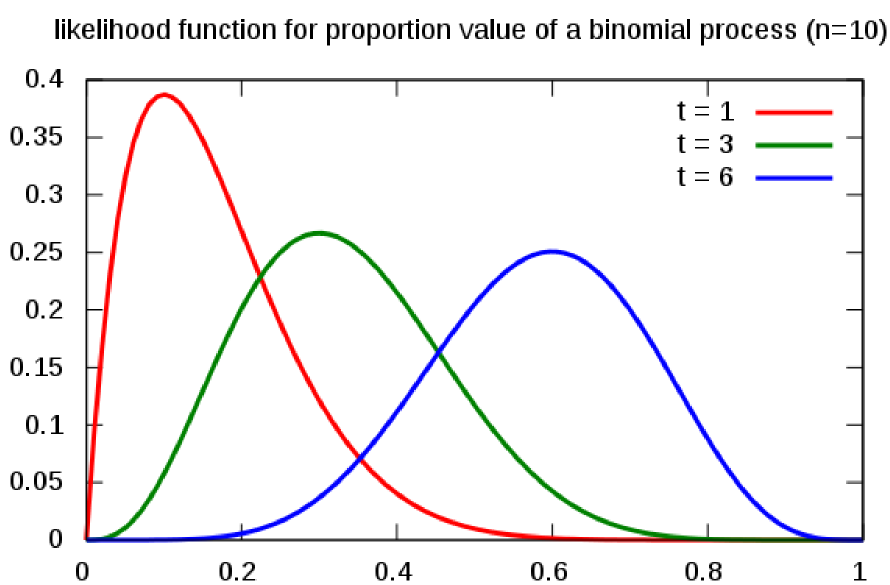


Figure 2: From `https://en.wikipedia.org`.

2.1 Normal distribution

Here, the parameter θ is (μ, σ). Observe that

$$\begin{aligned}\max_{\omega\in\Omega}\prod_{i=1}^{n} f_\omega(X_i) &= \max_{\omega\in\Omega}\left(\frac{1}{2\pi\sigma^2}\right)^{n/2}\exp\left(-\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{2\sigma^2}\right)\\ &= \max_{\omega\in\Omega} -(n/2)\log(2\pi\sigma^2) - (2\sigma^2)^{-1}\sum_{i=1}^{n}(X_i-\mu)^2\end{aligned}$$

Take the derivative of the objective function with respect to μ and setting it equal to zero, we get

$$0 + \frac{2n\sum_{i=1}^{n}(X_i-\mu)}{2\sigma^2} = 0,$$

or $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Next, take the derivative of the objective function with respect to σ and setting it equal to zero, we get

$$-n/\sigma + \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^3} = 0,$$

or $\hat{\sigma}_n = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$. Hence, The maximum likelihood estimator is not the same as the unbiased estimator.

3 References

- TOPE Chapter 2.
- Robert W. Keener's "Theoretical Statistics: Topics for a Core Course."