18.650. Statistics for Applications
Fall 2016. Problem Set 2

Due Friday, Sep. 23 at 12 noon

Problem 1 Biased/unbiased estimation

Let $X_1, \ldots, X_n$ be i.i.d. Bernoulli random variables, with unknown parameter $p \in (0,1)$. The aim of this exercise is to estimate the common variance of the X_i's.

1. Show that $\mathsf{var}(X_i) = p(1-p)$
2. Let $\bar{X}_n$ be the sample average of the X_i's. Prove that $\bar{X}_n(1-\bar{X}_n)$ is a consistent estimator of $p(1-p)$.
3. Compute the bias of this estimator.
4. Using the previous question, find an unbiased estimator of $p(1-p)$.

Problem 2 Statistical models and identifiability

For the following experiments, define a statistical model and check whether the parameter of interest is identified.

1. One observes n i.i.d. Poisson random variables with unknown parameter λ.
2. One observes n i.i.d. exponential random variables with parameter λ, which is unknown but a priori known to be no larger than 10.
3. One observes n i.i.d. uniform random variables in the interval $[0, \theta]$, where θ is unknown.
4. One observes n i.i.d. Gaussian random variables with unknown parameters μ, σ^2.
5. One observes the sign of n i.i.d. Gaussian random variables with unknown parameters μ, σ^2.
6. StatGen is a statistical procedure to test the relevance of genes. When well calibrated, it outputs the (random) proportion of active genes in a (random) cell. We want to estimate the distribution of this proportion. To that end, we take n iid cells and submit them to StatGen to collect n random variables $X_1, \ldots, X_n$ that have uniform distribution on $[0, \theta]$ for some unknown theta.
7. The US Census Bureau is interested in finding out the average commute time of Bostonians. To that end, it randomly selects n individuals, with replacement, among the people who work and live in the Boston area, and asks to each if their commute time is at least 20 minutes. The commute time of a random person is assumed to follow an exponential distribution.

8. Willy Wonka's contains 67 identical machines. Each machine has a lifetime that is modeled as an exponential random variable with some unknown parameter λ. After a certain time $T = 500$ days, one has observed the lifetime of all machines that have stopped working before T. The parameter of interest is λ.

Problem 3 A confidence interval for Poisson distributions

Let $X_1, \ldots, X_n$ be i.i.d. Poisson random variables with parameter $\lambda > 0$ and denote by $\bar{X}_n$ their empirical average:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

1. Find two sequences $(a_n)_{n \in \mathbb{N}^*}$ and $(b_n)_{n \in \mathbb{N}^*}$ such that $a_n(\bar{X}_n - b_n)$ converges in distribution to a standard Gaussian random variable $Z \sim N(0, 1)$.
2. Prove that for all $t > 0$,
$$\mathbb{P}[|Z| \leq t] = 2\mathbb{P}[Z \leq t] - 1.$$
3. Using the previous questions, find an interval $\mathcal{I}$ centered around $\bar{X}_n$ such that
$$\mathbb{P}[\mathcal{I} \ni \lambda] \to .95, \quad n \to \infty$$
`[Hint: The` 97.5%`-quantile of the standard Gaussian distribution is 1.96.]`
4. Modify the previous interval $\mathcal{I}$ in order to get a new interval $\mathcal{J}$, not necessarily centered around $\bar{X}_n$, that does not depend on λ and such that
$$\mathbb{P}[\mathcal{J} \ni \lambda] \to .95, \quad n \to \infty.$$

Problem 4 A confidence interval for uniform distributions

Let $X_1, \ldots, X_n$ be i.i.d. uniform random variables in $[0, \theta]$, for some $\theta > 0$. Denote by $M_n = \max_{i=1,\ldots,n} X_i$.

1. Prove that M_n converges in probability to θ.
2. Compute the cumulative distribution function of $n(1 - M_n/\theta)$ and prove that $n(1 - M_n/\theta)$ converges in distribution to an exponential random variable with parameter 1.
3. Use the previous question to find an interval $\mathcal{I}$ of the form $\mathcal{I} = [M_n, M_n + c]$, that does not depend on θ and such that
$$\mathbb{P}[\mathcal{I} \ni \theta] \to .95, \quad n \to \infty.$$
`[Hint:` $\theta \geq M_n$`]`
4. Is M_n unbiased?

18.650 / 18.6501 Statistics for Applications
Fall 2016