II Università di Roma, Tor Vergata
Dipartimento d'Ingegneria Civile e Ingegneria Informatica
LM in Ingegneria dell'Informazione e dell'Automazione
Complementi di Probabilità e Statistica - Advanced Statistics
Instructors: Roberto Monte & Massimo Regoli
Final Test - 2020-02-25 - Statistics

Problem 1 *Let X be a geometrically distributed real random variable with unknown success probability $p \in (0,1)$ representing the occurrence of the first success at the nth trial, where $n \in \mathbb{N}$, in binary random experiment. An investigator wants to estimate p on the basis of a simple random sample $X_1, \ldots, X_n$ of size n drawn from X.*

1. *Assume the investigator applies the method of moments. What is the estimator $\hat{p}_n^M$? Hint: it may be useful to recall that*
$$\sum_{n=1}^{\infty} q^n = \frac{q}{1-q}$$
and that, since $q \in (0,1)$, the formula
$$\frac{d}{dq}\sum_{n=1}^{\infty} q^n = \sum_{n=1}^{\infty} \frac{d}{dq} q^n$$
holds true.

2. *Is $\hat{p}_n^M$ unbiased? Is $\hat{p}_n^M$ consistent in probability?*

3. *Assume the investigator applies the likelihood methods. What is the estimator $\hat{p}_n^{LM}$? Hint: pay attention to writing the density function of X.*

4. *Given a realization $x_1, \ldots, x_{50}$ of a sample $X_1, \ldots, X_{50}$ of size 50 drawn from X such that $\sum_{k=1}^{50} x_k = 150$ apply the formulas obtained to get the realizations $\hat{p}_n^M(\omega)$ and $\hat{p}_n^{LM}(\omega)$ of the estimators $\hat{p}_n^M$ and $\hat{p}_n^{LM}$.*

5. *Give an estimate of the mean and the variance of X.*

Solution. .

Problem 2 *Let X a random variable representing a trait of a population. Assume that X has a density $f_X : \mathbb{R} \to \mathbb{R}$ given by*
$$f_X(x) \stackrel{def}{=} \frac{1}{\theta} e^{-\frac{x-3}{\theta}} 1_{[3,+\infty)}(x), \quad \forall x \in \mathbb{R}.$$

1. *Apply the method of moments to find the estimator $\hat{\theta}_{MM}$ of the parameter θ.*

2. *Apply the maximum likelihood method to find the estimator $\hat{\theta}_{ML}$ of the parameter θ.*

3. *Check whether $\hat{\theta}_n^M$ is unbiased and consistent in probability. Can you also check whether $\hat{\theta}_n^M$ is consistent in square mean?*

4. *Use the estimators $\hat{\theta}_{MM}$ and $\hat{\theta}_{ML}$ to build estimators for $\mathbf{E}[X]$ and $\mathbf{D}^2[X]$.*

Solution. .

Problem 3 *Assume that a library master believes that the mean duration in days of the borrowing period is* 20*d. However, the library master selects a simple random sample of* 100 *books in the library and discovers that the sample mean and standard deviation of the borrowing days are* 18*d and* 4*d, respectively.*

1. *Determine a* 95% *confidence interval for the mean duration of the borrowing days to check if library master's initial guess is correct.*
2. *Test the null hypothesis* $H_0 : \mu = 20$ *against the alternative* $H_1 : \mu > 20$ *at the significance level* $\alpha = 0.05$.
3. *Can you also compute the probability* $\beta(23)$ *of a II type error?*

Solution. .

Problem 4 *A random sample of* 25 *chickens from a hen-house are selected and two diets are given to them to increase their weight. A sub-sample of* 13 *chickens are fed with diet* X *and the remaining* 12 *chickens are fed with diet* Y. *After one month the chickens of the two samples are weighted and the weights in kg are the following*

Diet X	1.60,	1.70,	1.55,	1.70,	1.75,	1.35,	1.65,	1.75,	1.70,	1.65,	1.56,	1.60,	1.80.
Diet Y	1.70	1.35	1.50	1.65	1.50	1.60	1.70	1.75	1.60,	1.65,	1.55,	1.45.	

The weight of a chicken is assumed to be modeled by a normal random variable.

1. *Find the mean* μ_X *and* μ_Y *and the standard deviation* σ_X *and* σ_Y *for the weight of each sample.*
2. *Find an estimate of the pooled variance* s_p^2 *which estimates the variance of the whole sample.*
3. *Find a confidence interval for* $\mu_X - \mu_Y$ *at the* 95% *confidence level.*
4. *Formulate a null and an alternative hypothesis from the farmer's point of view who aims to test whether diet* X *is better than* Y. *Test such a null hypothesis against the alternative at the* 5% *significance level by using the rejection method and the p-value method.*

Solution. .

Problem 5 *Let* X *[resp.* Y*] be a Gaussian distributed random variables with (unknown) mean* $\mu_X \in \mathbb{R}$ *[resp.* $\mu_Y \in \mathbb{R}$*] and variance* $\sigma_X^2 > 0$ *[resp.* $\sigma_Y^2 > 0$*]. Assume that* X *and* Y *describe the same trait of different populations, for instance the cholesterol concentration in the blood of* 25+ *years aged males in Iceland and Democratic Republic of Congo in* 2008. *Then it is rather natural to assume that* X *and* Y *are idependent. Let* $X_1, \ldots, X_m$ *[resp.* $Y_1, \ldots, Y_n$*] be a simple random sample drawn by* X *[resp. by* Y*]. Assume to have measured*

$$\begin{array}{lllllllll} x_1 = 5.80 & x_2 = 5.70 & x_3 = 5.30 & x_4 = 5.60 & x_5 = 5.90 & x_6 = 6.20 & x_7 = 6.10 & & \\ y_1 = 3.20 & y_2 = 3.40 & y_3 = 4.70 & y_4 = 3.30 & y_5 = 4.50 & y_6 = 3.60 & y_7 = 4.30 & y_8 = 4.20 & y_9 = 4.10 \end{array}.$$

1. *Should we reject the null hypothesis* $H_0 = \mu_Y = \mu_X$ *against the alternatives* $H_1 = \mu_Y \neq \mu_X$ *and* $H_1 = \mu_Y > \mu_X$ *at the significance level* $\alpha = 0.05$*? Consider both the rejection region method and the p-value method.*

2. *Under the assumption that two random variables U and V are independent and such that $U \sim \chi^2_m$ and $V \sim \chi^2_n$, we know that the random variable $(U/m)/(V/n)$ has the Fisher-Snedecor distribution with m numerator degrees of freedom and n denominator degrees of freedom. In symbols,*

$$U \sim \chi^2_m,\ V \sim \chi^2_n,\ U \perp\!\!\!\perp V \Rightarrow \frac{U/m}{Y/n} \sim F(m,n).$$

We also know that the density of any Fisher-Snedecor distribution is concentrated on the positive real axis and is not symmetric (the graph of the density is similar to the graph of a χ^2 density). For some standard values of α (for instance $\alpha = 0.1, \alpha = 0.05, \alpha = 0.25, \alpha = 0.01, \dots$) the lower [resp. upper] critical value $f^-_{m,n,\alpha}$ [resp. $f^+_{m,n,\alpha}$] of $F(m,n)$ can be found in statistical tables or computed by almost all statistical softwares, for several values of m and n. Therefore, we assume to know $f^-_{m,n,\alpha}$ and $f^+_{m,n,\alpha}$. Now, in light of what we know, can you introduce a statistic to test the null hypothesis $H_0 : \sigma^2_X = \sigma^2_Y$ against the alternative $H_1 : \sigma^2_X > \sigma^2_Y$ and $H_1 : \sigma^2_X < \sigma^2_Y$ and describe for this statistic both the rejection method and the p-value method?

Solution. .