MSc. Finance & CLEFIN A.A. 2013/2014

Prep Course in Statistics – Prof. Massimo Guidolin

Review Questions concerning, Random Sampling and Point Estimation

SUGGESTION: try to approach the questions first, without looking at the answers. Occasionally the questions/problems explicitly send you to check out concepts and definitions that have not explicitly discussed during the prep course. In this case—besides Casella and Berger's book—Wikipedia may be very useful, as always.

1. (A case in which more data fail to mean higher precision of an estimate/sample statistic) Casella and Berger (i.e., your textbook, 2002, example 5.2.10) show that if Z_1, ..., Z_n is a random sample from a Cauchy$(0, 1)$ distribution, then the sample mean $\bar{Z}_n$ also has a Cauchy$(0, 1)$ distribution. Take that result as given (but make sure to read example 5.2.10). Show that if X_1, ..., X_n is a random sample from a Cauchy(μ, σ) distribution, then $\bar{X}_n$ also has a Cauchy(μ, σ) distribution. What do you learn from this result? [*Hint*: yes, now it is a good time to review what is a Cauchy distribution, see p. 107 of your textbook]

2. (Properties of the t-Student) Let X be a random variable with a t distribution with p degrees of freedom.

(a) Show that X^2 has an F distribution with 1 and p degrees of freedom, where Snedecor's F distribution with n and m degrees of freedom is the distribution of ratios of sample variances from two independent normal random samples, that we know have a chi-square distribution.[1] What is the link between the square of a t_p and $F_{1,p}$?

(b) Derive the mean and variance of X assuming $p > 2$. (*Hint*: exploit the fact that if and only if $X \sim t_p$, then $X = Z/\sqrt{V/p}$ where $Z \sim N(0, 1)$, $V \sim \chi^2_p$, and Z and V are independent).

(c) Let $f(x; p)$ denote the PDF of X. Show that

$$\lim_{p\to\infty} f(x;p) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

at each value of $x \in (-\infty, \infty)$. This suggests that as $p \to \infty$, X converges in distribution to a $N(0, 1)$ random variable. (*Hint*: Use Stirling's Formula, i.e., that $n! \simeq \sqrt{2\pi} n^{n+0.5} e^{-n}$, to make the proof easier; recall that if a_1, a_2, ... is a sequence of numbers converging to a, that is, $\lim_{n\to\infty} a_n = a$, then $\lim_{n\to\infty}(1 + a_n/(n + \kappa))^n = e^a$, where κ is a constant).

(d) Use the results of parts (a) and (b) to argue that, as $p \to \infty$, X^2 converges in distribution to a χ^2_1 random variable.

3. (The weak law of large numbers) State and, using Chebychev's Inequality, prove the weak law of large numbers. (*Hint*: Chebychev's Inequality states that if X is a random variable and $g(\cdot)$ is any nonnegative function, then, for any $r > 0$,

$$\Pr(g(X) \geq r) \leq \frac{E[g(X)]}{r}).$$

4. (Approximating the mean and variance of a ratio of random variables) Ratios of random variables are often very important in finance: think about accounting and focus on your favorite (if you really have one, I would start worrying) accounting ratio, defined as X/Y, where both X and Y are random variables, that are not known in advance (say X = current liabilities; Y = total assets). For simplicity, assume that both X and Y are continuous random variables and call their means μ_X and μ_Y, their variances σ^2_X, σ^2_Y, and their covariance σ_{XY}.

[1] The density of a $F_{n,m}$ distribution is: $f_F(x) = \frac{\Gamma\left(\frac{n+m}{2}\right)}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})} \left(\frac{n}{m}\right)^{\frac{n}{2}} \frac{x^{(n/2)-1}}{[1+(n/m)x]^{\frac{n+m}{2}}}$ $\quad t \in (0, +\infty)$.and can be shown to be ratio between a χ^2_n and a χ^2_m. Finally, $E[F_{n,m}] = m/(m-2)$ for $m > 2$ and $Var[F_{n,m}] = 2\left(\frac{m}{m-2}\right)^2 \frac{n+m-2}{n(m-4)}$ for $m > 4$.

(a) Argue (this is less than proving it) that

$$E\left[\frac{X}{Y}\right] \neq \frac{E[X]}{E[Y]} \qquad Var\left[\frac{X}{Y}\right] \neq \frac{Var[X]}{Var[Y]}.$$

(b) Using a first-order Taylor expansion around the point $[\mu_X \ \mu_Y]'$, show that as long as the conditions of Taylor's theorem hold,

$$\hat{E}\left[\frac{X}{Y}\right] \simeq \frac{\bar{X}}{\bar{Y}} \qquad \widehat{Var}\left[\frac{X}{Y}\right] \simeq \left(\frac{\bar{X}}{\bar{Y}}\right)^2 \left(\frac{S_X^2}{\bar{X}^2} + \frac{S_Y^2}{\bar{Y}^2} - 2\frac{S_{XY}}{\bar{X}\bar{Y}}\right),$$

where $\bar{X} \equiv n^{-1}\sum_{i=1}^n X_i$, $S_X^2 \equiv (n-1)^{-1}\sum_{i=1}^n (X_i - \bar{X})^2$, $S_{XY} \equiv (n-2)^{-1}\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})^2$, and the last sample statistic is a sample covariance. Here "hats" indicate estimated moments.

5. (Higher-order moments for normal random samples) Let X_1, ..., X_n be a random sample from a $N(\mu, \sigma^2)$ population.

(a) Find expressions for $\mu_1 \equiv E[X_i]$, $\mu_2 \equiv E[(X_i - \mu_1)^2]$, $\mu_3 \equiv E[(X_i - \mu_1)^3]$, and , $\mu_4 \equiv E[(X_i - \mu_1)^4]$ in terms of μ and σ^2. (*Hint*: in the case of μ_3, use Stein's lemma, that $E[g(X)(X - \mu)] = \sigma^2 E[g'(X)]$).

(b) Calculate $Var[S^2]$ using the fact that $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$ (stated in the lectures) and that $Var[\chi^2_{n-1}] = 2(n-1)$.

6. (The CLT and a naive confidence interval for the difference of sample means from independent random samples) If $\bar{X}_1$ and $\bar{X}_2$ are the means of two independent samples of size n from a population with variance σ^2, find a value for n so that $\Pr(|\bar{X}_1 - \bar{X}_2| < \sigma/5) \simeq 0.99$. Justify your calculations.

7. (Method of Moment Estimation in the Binomial Case) Let X_1, X_2, ..., X_n be IID $Bi(k, p)$, i.e.,

$$\Pr(X_i = x; k, p) = \left(\begin{array}{c} k \\ p \end{array}\right) p^x (1-p)^{k-x}, \ x = 0, 1, ..., \ k.$$

Find the method of moment estimators for k and p simultaneously. Can you obtain negatives estimates of the parameters? What can you learn about the features of the random sample when k and p are estimated to be negative by the method of moments?

8. (MLE Estimation of Mean and Variance from a Normal Population) Let X_1, ..., X_n be a random sample from a $N(\mu, \sigma^2)$ population with both μ and σ^2 unknown.

a. Derive the ML estimators of μ and σ^2 by writing and differentiating the log-likelihood function.

b. Verify that the estimators derived under (a) represent an interior maximum and not a minimum.

9. (The relative MSEs of sample variance vs. ML estimator for the variance) Let X_1, ..., X_n be a random sample from a $N(\mu, \sigma^2)$ population with both μ and σ^2 unknown.

a. Write the sample variance estimator, S^2, and the ML estimator for variance.

b. Compute the bias and the MSE of both estimators.

c. Show that even though it is biased, the ML estimator has a lower MSE than the sample variance.

10. (S^2 is not UMVUE) Let X_1, ..., X_n be a random sample from a $N(\mu, \sigma^2)$ population with both μ and σ^2 unknown and consider the sample variance estimator for σ^2. Show that S^2 is unbiased but does not reach the Cramer-Rao lower bound. [*Hint*: because the normal distribution belongs to the normal family, one can prove that

$$E\left\{\left[\frac{\partial}{\partial \boldsymbol{\theta}'} \ln f(\mathbf{x}; \boldsymbol{\theta})\right]^2\right\} = -nE\left\{\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \ln f(\mathbf{x}; \boldsymbol{\theta})\right\})$$

11. (Uniqueness of the Best Unbiased Estimator) Show that if W is BUE for θ, then W must be unique, i.e., if there is another W^* that is also claimed to be BUE, then $W^* = W$. (*Hint*: recall that the U in BUE means unbiased and that as such $MSE(W) = Var[W]$; then try to build a third estimator just by taking $0.5W + 0.5W^*$ and examine its MSE; finally, it could be useful to use the Cauchy-Schwarz inequality, by which $Cov[X, Y] \leq \sqrt{Var[X]Var[Y]}$).

12. (Consistency of the normal sample mean) Let X_1, ..., X_n be an IID sample from a $N(\mu, 1)$ population. Establish analytically that the sample mean estimator sequence $\bar{X}_n \equiv n^{-1}\sum_{i=1}^n X_i$ is consistent.