Università Commerciale Luigi Bocconi

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

# Prep Course in Statistics – Prof. Massimo Guidolin

## A Few Review Questions and Problems 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$(\mu,\sigma)$ distribution, then $\bar{X}_n$ also has a Cauchy$(\mu,\sigma)$ 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]

**Solution**. The Cauchy belongs to the location-scale family, in the sense that if $Z_1$, ..., $Z_n$ is a random sample from a Cauchy$(0,1)$ distribution and one sets

$$X_i = \mu + \sigma Z_i \qquad i = 1, 2, ..., N,$$

then $X_1$, ..., $X_n$ is a random sample from a Cauchy$(\mu,\sigma^2)$ distribution. Therefore, using the Key Result 2 in lecture 1, we know that

$$f(\bar{X}) = \frac{1}{\sigma} g(\bar{Z}) = \frac{1}{\sigma}\left(\frac{1}{\pi}\frac{1}{1+\left(\frac{\bar{Z}-\mu}{\sigma}\right)^2}\right) = \frac{1}{\pi\sigma}\frac{1}{1+\left(\frac{\bar{Z}-\mu}{\sigma}\right)^2},$$

which is the density of a Cauchy$(\mu,\sigma)$.

*The lesson of this exercise is:* In finance, we normally rely on the abundance of data on prices and returns, to say that most of our estimates—starting from the sample mean—will make very (increasingly) precise as we acquire more data. We tend to treat this principle as an obvious, universal truth. This exercise shows that this is not an obvious fact, but that this depends instead on the underlying distributional assumptions on the distribution. While Key Result 2 in lecture 1 shows that when the random sample comes from a normal distribution, then

$$\bar{X}_n^{Normal} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

so that $\lim_{n\to\infty}\sqrt{Var[\bar{X}_n^{Normal}]} = \lim_{n\to\infty}\sqrt{\sigma^2/n} = \lim_{n\to\infty}\sigma/\sqrt{n} = 0$, this exercise faces with a case in which

$$\bar{X}_n^{Cauchy} \sim Cauchy\,(\mu, \sigma)$$

so that the $\sigma$ parameter that enters the location-scale structure $X_i = \mu + \sigma Z_i$ is not divided by $n$: the precision of a Cauchy sample mean remains unchanged.[1]

**2. (Properties of the t-Student)** Let $X$ be a random variable with a Student's $t$ 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.[2] 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 $\kappa$ is a constant).

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

**Solution**. (a) Using the same hint as in (b), $X^2 = Z^2/(V/p)$. From lecture 1, we know $Z^2 \sim \chi^2_1$ and $V/p \sim \chi^2_p$. Therefore $X^2$ is the ration between a $\chi^2_1$ and a $\chi^2_p$ that according to the question, has a $F_{1,p}$ distribution. We know from lecture 1 that $X = Z/\sqrt{V/p} \sim t_p$ so that $X^2 \sim t^2_p$ which establishes that $t^2_p \sim F_{1,p}$.

(b) Using the hint, and in particular the fact that $E[h(Z)g(V)] = E[h(Z)]E[g(V)]$ if $Z$ and $V$ are independent, then

$$E[X] = E\left[\frac{Z}{\sqrt{V/p}}\right] = E[Z]E\left[\frac{1}{\sqrt{V/p}}\right] = 0 \times E\left[\frac{1}{\sqrt{V/p}}\right] = 0$$

[1] But in the Cauchy case, the sample mean is no estimator of the expectation, the population mean. As you should recall from your Statistics courses, a Cauchy distribution has no finite moments (equivalently, its MGF does not exist) and such all such moments cannot be estimated. In case you are wondering, $\mu$ can be shown to be the median of the distribution.

[2] The density of a $F_{n,m}$ distribution is:

$$f_F(x) = \frac{\Gamma\left(\frac{n+m}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)}\left(\frac{n}{m}\right)^{\frac{n}{2}}\frac{x^{(n/2)-1}}{[1+(n/m)x]^{\frac{n+m}{2}}} \qquad t \in (0,+\infty).$$

and can be shown to be ratio between a $\chi^2_n$ and a $\chi^2_m$. Finally, $E[F_{n,m}] = m/(m-2)$ for $m > 2$ and

$$Var[F_{n,m}] = 2\left(\frac{m}{m-2}\right)^2\frac{n+m-2}{n(m-4)} \text{ for } m > 4.$$

as long as the other expectation is finite. From the properties of a chi-square distribution, this is so if $p > 1$. From part (a), $X^2 \sim F_{1,p}$, therefore

$$Var[X] = E[X^2] - \underbrace{\{E[X]\}^2}_{=0} = E[X^2] = \frac{p}{p-2} < \infty$$

for $p > 2$.

(c) The pdf of $X$ is

$$f_X(x) = \frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)} \frac{1}{\sqrt{p\pi}} \frac{1}{\left(1 + \frac{x^2}{p}\right)^{\frac{p+1}{2}}}.$$

Note that from $\Gamma(n) = (n-1)!$, we have

$$\begin{aligned}
\lim_{p\to\infty} \frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)} \frac{1}{\sqrt{p\pi}} &= \lim_{p\to\infty} \frac{\left(\frac{p-1}{2}\right)!}{\left(\frac{p-2}{2}\right)!} \frac{1}{\sqrt{p\pi}} \simeq \lim_{p\to\infty} \frac{\sqrt{2\pi}\left(\frac{p-1}{2}\right)^{\frac{p}{2}} e^{-\left(\frac{p-1}{2}\right)}}{\sqrt{2\pi}\left(\frac{p-2}{2}\right)^{\left(\frac{p-1}{2}\right)} e^{-\left(\frac{p-2}{2}\right)}} \frac{1}{\sqrt{p\pi}} \\
&= \frac{e^{-\frac{1}{2}}}{\sqrt{\pi}} \lim_{p\to\infty} \frac{\left(\frac{p-1}{2}\right)^{\frac{p-1}{2}+\frac{1}{2}}}{\left(\frac{p-2}{2}\right)^{\frac{p-2}{2}+\frac{1}{2}}} \frac{1}{\sqrt{p}} = \frac{e^{-\frac{1}{2}}}{\sqrt{\pi}} \frac{e^{\frac{1}{2}}}{\sqrt{2}} = \frac{1}{\sqrt{2\pi}}
\end{aligned}$$

while

$$\begin{aligned}
\lim_{p\to\infty} \left(1 + \frac{x^2}{p}\right)^{\frac{p}{2}+\frac{1}{2}} &= \lim_{p\to\infty} \left(1 + \frac{1}{2}\frac{2x^2}{p}\right)^{\frac{p}{2}+\frac{1}{2}} = \lim_{y\to\infty} \left(1 + \frac{1}{2}\frac{x^2}{y}\right)^{y+\frac{1}{2}} \quad (\text{where } y \equiv \frac{p}{2}) \\
&= \lim_{w\to\infty} \left(1 + \frac{(x^2/2)}{w - \frac{1}{2}}\right)^{w} = e^{x^2/2} \ (\text{where } w \equiv y + \frac{1}{2}),
\end{aligned}$$

(trust me, it is correct, I have checked using Mathematica) so that

$$\lim_{p\to\infty} \frac{1}{\left(1 + \frac{x^2}{p}\right)^{\frac{p+1}{2}}} = \frac{1}{\lim_{p\to\infty}\left(1 + \frac{x^2}{p}\right)^{\frac{p}{2}+\frac{1}{2}}} = e^{-x^2/2}.$$

Pulling these results together,

$$\lim_{p\to\infty} f_X(x) = \lim_{p\to\infty} \frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)} \frac{1}{\sqrt{p\pi}} \frac{1}{\left(1 + \frac{x^2}{p}\right)^{\frac{p+1}{2}}} \simeq \frac{1}{\sqrt{2\pi}} e^{-x^2/2},$$

which is obviously the density of a $N(0,1)$ random variable.

(d) This is a sensible statement, although the math is rather involved: as the random variable $F_{1,p}$ is the square of a $t_p$, we conjecture that it would converge to the square of a $N(0,1)$ random variable, a $\chi_1^2$. This means that as $p \to \infty$, $F_{1,p} \xrightarrow{D} \chi_1^2$, where $\xrightarrow{D}$ denotes converge in distribution.

*The lesson of this exercise is:* Because many estimators in finance (think of the sample mean, as seen in lecture 1) have a $t_{n-k}$ distribution, where $n$ is the sample size, when $n \to \infty$ one loses any advantage in distinguishing between the t-Student and a standard normal distribution with density $\phi(x) = (2\pi)^{-1/2} e^{-x^2/2}$. Moreover, while (a) and (d) are just math games that help us establish links between the distributional objects under investigation—let's recap here, normal distributions, t-Student, chi-square, Snedecor's F—(b) is helpful to emphasize that with t-Student distributions, variance exists only if there are at least more than 2 degrees of freedom, which is a very important point in finance.

**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 $r > 0$,

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

**Solution**. Applying Chebychev's Inequality to the probability in the WLLN statement,

$$\lim_{n\to\infty} \Pr(|\bar{X}_n - \mu| \geq \epsilon) = 0,$$

we have

$$\Pr(|\bar{X}_n - \mu| \geq \epsilon) = \Pr((\bar{X}_n - \mu)^2 \geq \epsilon^2) \leq \frac{E[(\bar{X}_n - \mu)^2]}{\epsilon^2} = \frac{Var[\bar{X}_n]}{\epsilon^2} = \frac{\sigma^2/n}{\epsilon^2}.$$

Hence

$$\begin{aligned} \lim_{n\to\infty} \Pr(|\bar{X}_n - \mu| &< \epsilon) = 1 - \lim_{n\to\infty} \Pr(|\bar{X}_n - \mu| \geq \epsilon) \\ &= 1 - \lim_{n\to\infty} \frac{\sigma^2}{n\epsilon^2} = 1. \end{aligned}$$

**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 worring) 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 $\mu_X$ and $\mu_Y$, their variances $\sigma_X^2$, $\sigma_Y^2$, and their covariance $\sigma_{XY}$.

(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.

**Solution**. (a) Call $f_{XY}(x, y)$ the continuous, joint PDF of $X$ and $Y$. Then

$$E\left[\frac{X}{Y}\right] = \int_{\text{supp}(X)} \int_{\text{supp}(Y)} \frac{x}{y} f_{XY}(x, y) dx dy,$$

where supp$(\cdot)$ indicates the support of a continuous random variable. Clearly,

$$\int_{\text{supp}(X)} \int_{\text{supp}(Y)} \frac{x}{y} f_{XY}(x, y) dx dy \neq \frac{\int_{\text{supp}(X)} x f_X(x) dx}{\int_{\text{supp}(Y)} y f_Y(y) dy} = \frac{\int_{\text{supp}(X)} x \left[\int_{\text{supp}(Y)} f_{XY}(x, y) dy\right] dx}{\int_{\text{supp}(Y)} y \left[\int_{\text{supp}(X)} f_{XY}(x, y) dx\right] dy}.$$

As for $Var\left[\frac{X}{Y}\right]$, note that

$$Var\left[\frac{X}{Y}\right] = E\left[\left(\frac{X}{Y}\right)^2\right] - \left\{E\left[\frac{X}{Y}\right]\right\}^2$$

while

$$\frac{Var[X]}{Var[Y]} = \frac{E[X^2] - \{E[X]\}^2}{E[Y^2] - \{E[Y]\}^2}.$$

Clearly, $E\left[X/Y\right]$ has nothing to do with $E[X]/E[Y]$ and similarly, $E\left[(X/Y)^2\right]$ has nothing to do with $E[X^2]/E[Y^2]$. Heuristically at least, it should convince yourself that also the variance of a ratio is hardly the ratio of the variances.

(b) Using a first-order Taylor expansion of the function $g(X, Y) = X/Y$ of vector of random variables $[X\ Y]'$ around $[\mu_X\ \mu_Y]'$, we obtain:

$$\begin{aligned}
\frac{x}{y} &= g(x,y) \simeq \frac{\mu_X}{\mu_Y} + \left.\frac{d(x/y)}{dx}\right|_{x=\mu_X,\ y=\mu_Y}(x-\mu_X) + \left.\frac{d(x/y)}{dy}\right|_{x=\mu_X,\ y=\mu_Y}(y-\mu_Y) \\
&= \frac{\mu_X}{\mu_Y} + \frac{1}{\mu_Y}(x-\mu_X) - \frac{\mu_X}{\mu_Y^2}(y-\mu_Y).
\end{aligned}$$

At this point, simple rules on how to compute expectations and variances, give:

$$\begin{aligned}
E\left[\frac{X}{Y}\right] &\simeq \frac{\mu_X}{\mu_Y} \qquad Var\left[\frac{X}{Y}\right] \simeq \frac{1}{\mu_Y^2}\sigma_X^2 + \frac{\mu_X^2}{\mu_Y^4}\sigma_Y^2 - 2\frac{\mu_X}{\mu_Y^3}\sigma_{XY} \\
&= \frac{\mu_X^2}{\mu_Y^2}\left(\frac{\sigma_X^2}{\mu_X^2} + \frac{\sigma_Y^2}{\mu_Y^2} - 2\frac{\sigma_{XY}}{\mu_X\mu_Y}\right)
\end{aligned}$$

At this point, as sample means are "good" estimates for $\mu_X$ and $\mu_Y$ and sample variances are "good" estimates for $\sigma_X^2$ and $\sigma_Y^2$ (and you can easily imagine that the same applies to sample covariance with reference to $\sigma_{XY}$), by plugging in we have:

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

*The lesson of this exercise is:* Let's put it lightly, your head should be chopped off every time you toy with the idea of setting $E\left[X/Y\right] \neq E[X]/E[Y]$ and yet, that would be too harsh because, up to the technicalities of a first-order Taylor approximation, we have shown that $E\left[X/Y\right] \simeq \bar{X}/\bar{Y}$. Note that contrary to common belief, mathematics is not the enemy of simple short-cuts—instead it just helps you understand under what conditions (if any) these are viable. However, there is no way you can ever defend the tragic thought that $Var\left[X/Y\right] = Var[X]/Var[Y]$, which is plainly WRONG.

**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 $\mu$ and $\sigma^2$. (*Hint*: in the case of $\mu_3$, 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 $Var[\chi^2_{n-1}] = 2(n-1)$.

**Solution**. (a) Trivially, $\mu_1 \equiv E[X_i] = \mu$ and $\mu_2 \equiv E[(X_i - \mu_1)^2] = \sigma^2$ by definition. The most interesting central moments are instead:

$$\begin{aligned}
\mu_3 &\equiv E[(X_i-\mu_1)^3] = E[(X_i-\mu_1)^2(X_i-\mu_1)] = \underbrace{\sigma^2 E\left[\frac{d[(X_i-\mu_1)^2]}{dX_i}\right]}_{\text{by Stein's lemma}} \\
&= 2\sigma^2 E[(X_i-\mu_1)] = 0 \text{ (as } E[X_i]=\mu_1) \\
\mu_4 &\equiv E[(X_i-\mu_1)^4] = E[(X_i-\mu_1)^3(X_i-\mu_1)] = \sigma^2 E\left[\frac{d[(X_i-\mu_1)^3]}{dX_i}\right] \\
&= 3\sigma^2 E[(X_i-\mu_1)^2] = 3\sigma^2\cdot\sigma^2 = 3\sigma^4
\end{aligned}$$

(b)

$$\begin{aligned}
Var[S^2] &= Var\left[\frac{\sigma^2}{(n-1)}\chi^2_{n-1}\right] = \frac{\sigma^4}{(n-1)^2}Var\left[\chi^2_{n-1}\right] \\
&= \frac{2\sigma^4}{(n-1)^2}(n-1) = \frac{2\sigma^4}{n-1}.
\end{aligned}$$

*The lesson of this exercise is:* The fact that the variance of the sample variance (yes, this is what $Var[S^2]$ is) is $2\sigma^4/(n-1)$ is often useful in finance. Notice that $2\sigma^4/(n-1) = (2/3)\mu_4/(n-1)$, which is a scaled version of $\mu_4$, the fourth central moment of a normal distribution. Note that in the same way in which $Var[\bar{X}] = \sigma^2/n \to 0$ as $n \to \infty$, $Var[S^2] \to 0$ as $n \to \infty$, even though the latter quantity depends on $\sigma^4$, not $\sigma^2$. $\mu_3 = 0$ is interpreted to mean that normal random samples are symmetric around the mean, as all of their odd central (i.e., around the mean) moments are zero.

**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 $\sigma^2$, find a value for $n$ so that $\Pr(|\bar{X}_1 - \bar{X}_2| < \sigma/5) \simeq 0.99$. Justify your calculations.

**Solution**. From the CLT we have, approximately, $\bar{X}_1 \sim N(\mu, \sigma^2/n)$, $\bar{X}_1 \sim N(\mu, \sigma^2/n)$. Because $\bar{X}_1$ and $\bar{X}_2$ are independent, $\bar{X}_1 - \bar{X}_2 \sim N(\mu, 2\sigma^2/n)$ as

$$\begin{aligned}
Var[\bar{X}_1 - \bar{X}_2] &= Var[\bar{X}_1] + (-1)^2 Var[\bar{X}_2] - 2Cov[\bar{X}_1, \bar{X}_2] \\
&= \frac{\sigma^2}{n} + \frac{\sigma^2}{n} - 2\times 0 = \frac{2\sigma^2}{n} = \frac{\sigma^2}{n/2}.
\end{aligned}$$

Thus, we want

$$\begin{aligned}
0.99 &\simeq \Pr(|\bar{X}_1 - \bar{X}_2| < \sigma/5) = \Pr\left(\frac{|\bar{X}_1 - \bar{X}_2|}{\sigma/\sqrt{n/2}} < \frac{\sigma/5}{\sigma/\sqrt{n/2}}\right) \\
&= \Pr\left(-\frac{1/5}{1/\sqrt{n/2}} < \underbrace{\frac{\bar{X}_1 - \bar{X}_2}{\sigma/\sqrt{n/2}}}_{Z\sim N(0,1)} < \frac{1/5}{1/\sqrt{n/2}}\right) = \Pr\left(-\frac{\sqrt{n}}{5\sqrt{2}} < Z < \frac{\sqrt{n}}{5\sqrt{2}}\right).
\end{aligned}$$

Because

$$1 - 0.99 = 0.01 \simeq 1 - \Pr\left(-\frac{\sqrt{n}}{5\sqrt{2}} < Z < \frac{\sqrt{n}}{5\sqrt{2}}\right),$$

we need $\Pr\left(Z \geq \frac{\sqrt{n}}{5\sqrt{2}}\right) = \frac{0.01}{2} = 0.005$. From standard normal tables, the corresponding critical value is 2.576 and

$$\frac{\sqrt{n}}{5\sqrt{2}} = 2.576 \Longrightarrow n^* = (2.576)^2 25 \times 2 \simeq 332.$$

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

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

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?

**Solution**. Recalling that in the case of a binomial, $E[X] = kp$ and $Var[X] = kp(1-p)$, then

$$\begin{aligned}
m_1 &= n^{-1}\sum_{i=1}^{n} X_i = kp \Longrightarrow \hat{p} = \frac{\bar{X}}{\hat{k}} \\
m_2 &= n^{-1}\sum_{i=1}^{n} X_i^2 = kp(1-p) + kp \Longrightarrow n^{-1}\sum_{i=1}^{n} X_i^2 = \hat{k}\hat{p}(1-\hat{p}) + \hat{k}\hat{p} = \hat{k}\hat{p}(2-\hat{p}) \\
&\Longrightarrow n^{-1}\sum_{i=1}^{n} X_i^2 = \hat{k}\frac{\bar{X}}{\hat{k}}\left(2 - \frac{\bar{X}}{\hat{k}}\right) = 2\bar{X} - \frac{\bar{X}^2}{\hat{k}} \Longrightarrow \left(2\bar{X} - n^{-1}\sum_{i=1}^{n} X_i^2\right) = \frac{\bar{X}^2}{\hat{k}} \\
&\Longrightarrow \hat{k} = \frac{\bar{X}^2}{2\bar{X} - n^{-1}\sum_{i=1}^{n} X_i^2} = \frac{\bar{X}^2}{\bar{X} - n^{-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}.
\end{aligned}$$

At this point, plugging the expression for $\hat{k}$ back into the expression for $\hat{p}$, we obtain:

$$\hat{p} = \bar{X}\frac{\bar{X} - n^{-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}{\bar{X}^2} = \frac{\bar{X} - n^{-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}{\bar{X}} = 1 - \frac{n^{-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}{\bar{X}}.$$

Admittedly, these are not the best estimators for the population parameters. In particular, from a few possible random samples, it may happen that one obtained negative estimates of $k$ and $p$ which, of course, cannot be. This is a case where the range of the estimator does not coincide with the range of the parameter you are estimating. However, for $\hat{k}$ or $\hat{p}$ to obtain, one needs

$$\bar{X} - n^{-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 < 0 \text{ or } \bar{X} < n^{-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{n-1}{n}S^2.$$

This means that the sample mean is smaller than the sample variance, indicating a large degree of variability in the data.

**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 $\mu$ and $\sigma^2$ unknown.

a. Derive the ML estimators of $\mu$ and $\sigma^2$ by writing and differentiating the log-likelihood function.

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

**Solution**. (a) The log-likelihood is:

$$\begin{aligned}
\ln L(\mu,\sigma^2;\mathbf{x}) &= \ln\left\{\textstyle\prod_{i=1}^n \frac{1}{(2\pi\sigma^2)^{1/2}}\exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right]\right\} \\
&= \ln\frac{1}{(2\pi\sigma^2)^{n/2}} - \sum_{i=1}^n \frac{(x_i-\mu)^2}{2\sigma^2} \\
&= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2}\sum_{i=1}^n \frac{(x_i-\mu)^2}{\sigma^2}.
\end{aligned}$$

The partial derivatives with respect to $\mu$ and $\sigma^2$ are:

$$\begin{aligned}
\frac{\partial L(\mu,\sigma^2;\mathbf{x})}{\partial\mu} &= \frac{1}{\hat{\sigma}^2}\sum_{i=1}^n (x_i-\hat{\mu}) = 0 \Longrightarrow \hat{\mu} = \frac{1}{n}\sum_{i=1}^n x_i = \bar{X}_n \\
\frac{\partial L(\mu,\sigma^2;\mathbf{x})}{\partial\sigma^2} &= -\frac{n}{2}\frac{1}{\hat{\sigma}^2} + \frac{1}{2}\frac{1}{\hat{\sigma}^4}\sum_{i=1}^n (x_i-\hat{\mu})^2 \\
&= -\frac{n}{2}\frac{1}{\hat{\sigma}^2} + \frac{1}{2}\frac{1}{\hat{\sigma}^4}\sum_{i=1}^n (x_i-\bar{X})^2 = 0.
\end{aligned}$$

Solving this last equation, yields

$$\frac{n}{\hat{\sigma}^2} = \frac{1}{\hat{\sigma}^4}\sum_{i=1}^n (x_i-\bar{X})^2 \Longrightarrow \frac{1}{\hat{\sigma}^2}\sum_{i=1}^n (x_i-\bar{X})^2 = n \Longrightarrow \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (x_i-\bar{X})^2 = \frac{n-1}{n}S^2,$$

which shows that the MLE for the variance is not identical to the sample variance, but it is a $(n-1)/n$ factor of the same.

(b) In this case, one has to verify that the matrix of second-order partial derivatives of $L(\mu,\sigma^2;\mathbf{x})$ with respect to $\mu$ and $\sigma^2$ (collected in a $2\times 1$ vector $\boldsymbol{\theta}$), also called the Hessian of the problem, is definite negative in correspondence to the ML estimates derived under (a). The Hessian matrix is (recall that $\partial L^2(\mu,\sigma^2;\mathbf{x})/\partial\mu\partial\sigma^2 = \partial L^2(\mu,\sigma^2;\mathbf{x})/\partial\sigma^2\partial\mu$):

$$\begin{aligned}
H(\hat{\boldsymbol{\theta}}^{MLE};\mathbf{x}) &\equiv \left.\frac{\partial^2 L(\boldsymbol{\theta};\mathbf{x})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \begin{bmatrix} \left.\frac{\partial L^2(\mu,\sigma^2;\mathbf{x})}{\partial\mu^2}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} & \left.\frac{\partial L^2(\mu,\sigma^2;\mathbf{x})}{\partial\mu\partial\sigma^2}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \\ \left.\frac{\partial L^2(\mu,\sigma^2;\mathbf{x})}{\partial\sigma^2\partial\mu}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} & \left.\frac{\partial L^2(\mu,\sigma^2;\mathbf{x})}{\partial(\sigma^2)^2}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \end{bmatrix} \\
&= \begin{bmatrix} -\frac{n}{\hat{\sigma}^2} & -\frac{1}{\hat{\sigma}^4}\sum_{i=1}^n (x_i-\hat{\mu}) \\ -\frac{1}{\hat{\sigma}^4}\sum_{i=1}^n (x_i-\hat{\mu}) & \frac{n}{2}\frac{1}{\hat{\sigma}^4} - \frac{1}{\hat{\sigma}^6}\sum_{i=1}^n (x_i-\hat{\mu})^2 \end{bmatrix} \\
&= \begin{bmatrix} -\frac{n}{\hat{\sigma}^2} & 0 \\ 0 & \frac{n}{2}\frac{1}{\hat{\sigma}^4} - \frac{1}{\hat{\sigma}^6}\sum_{i=1}^n (x_i-\hat{\mu})^2 \end{bmatrix}
\end{aligned}$$

The determinant of the Hessian is

$$\begin{aligned}
-\frac{n}{\hat{\sigma}^2}\left[\frac{n}{2}\frac{1}{\hat{\sigma}^4} - \frac{1}{\hat{\sigma}^6}\sum_{i=1}^n (x_i-\hat{\mu})^2\right] &= -\frac{n}{\hat{\sigma}^2}\left[\frac{n}{2}\frac{1}{\hat{\sigma}^4} - \frac{n}{\hat{\sigma}^4}\right] \\
&= \frac{n^2}{\hat{\sigma}^6}\left[\frac{1}{2}-1\right] = -\frac{1}{2}\frac{n^2}{\hat{\sigma}^6} < 0,
\end{aligned}$$

while $\partial L^2(\mu,\sigma^2;\mathbf{x})/\partial\mu^2\big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} < 0$ and

$$\left.\frac{\partial L^2(\mu,\sigma^2;\mathbf{x})}{\partial(\sigma^2)^2}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \frac{n}{2}\frac{1}{\hat{\sigma}^4} - \frac{1}{\hat{\sigma}^6}\sum_{i=1}^n (x_i-\hat{\mu})^2 = \frac{n}{2}\frac{1}{\hat{\sigma}^4} - \frac{n}{\hat{\sigma}^4} = -\frac{1}{2}\frac{n}{\hat{\sigma}^4} < 0.$$

These conditions establish that the Hessian is definite negative and as a result, $\hat{\boldsymbol{\theta}}^{MLE}$ computed as in (a) represents an interiour maximum for the log-likelihood function.

*The lesson of this exercise is:* Once more, the sample mean emerges as the usual suspect estimator for the mean, because as we saw it is both the method of moments as well as the ML estimator. On the contrary, some small differences between the ML estimator of the variance and the sample variance have emerged, as $\hat{\sigma}^2 = n^{-1}\sum_{i=1}^n (x_i - \bar{X})^2 = (n - 1/n)S^2$.

**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 $\mu$ and $\sigma^2$ 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.

**Solution**. (a) By simply looking at the class lectures and at previous exercises, we know that

$$S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2 \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2 = \frac{n-1}{n}S^2.$$

(b) Because we know that $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$, therefore

$$E[S^2] = E\left[\frac{\sigma^2}{n-1}\chi^2_{n-1}\right] = \frac{\sigma^2}{n-1}E[\chi^2_{n-1}] = \sigma^2.$$

Moreover,

$$E[\hat{\sigma}^2] = E\left[\frac{n-1}{n}S^2\right] = \frac{n-1}{n}E\left[S^2\right] = \frac{n-1}{n}\sigma^2$$

and this implies that while $S^2$ is unbiased, $\hat{\sigma}^2$ is not; such as bias is

$$bias(\hat{\sigma}^2) = E[\hat{\sigma}^2] - \sigma^2 = \frac{n-1}{n}\sigma^2 - \sigma^2 = \left[\frac{n-1}{n} - 1\right]\sigma^2 = -\frac{1}{n}\sigma^2$$

Therefore:

$$MSE[S^2] = Var[S^2] = \frac{2\sigma^4}{n-1}$$

as seen in problem 5b;

$$\begin{aligned} MSE[\hat{\sigma}^2] &= Var[\hat{\sigma}^2] + \left[-\frac{1}{n}\sigma^2\right]^2 = Var\left[\frac{n-1}{n}S^2\right] + \left[-\frac{1}{n}\sigma^2\right]^2 \\ &= \frac{(n-1)^2}{n^2}Var\left[S^2\right] + \frac{1}{n^2}\sigma^4 = \frac{(n-1)^2}{n^2}\frac{2\sigma^4}{n-1} + \frac{1}{n^2}\sigma^4 \\ &= \frac{n-1}{n^2}2\sigma^4 + \frac{1}{n^2}\sigma^4 = \frac{\sigma^4}{n^2}\left[2n-2+1\right] = \frac{(2n-1)\sigma^4}{n^2}. \end{aligned}$$

(c) Finally, comparing the two MSEs, we have:

$$\begin{aligned} MSE[\hat{\sigma}^2] &= \frac{(2n-1)\sigma^4}{n^2} = \left(\frac{n-1}{n}\frac{2}{n-1} - \frac{1}{n^2}\right)\sigma^4 \\ &< \frac{n-1}{n}\frac{2}{n-1}\sigma^4 < \frac{2}{n-1}\sigma^4 = MSE[S^2], \end{aligned}$$

which shows that $MSE[\hat{\sigma}^2]$ has the lowest MSE.

*The lesson of this exercise is:* Very important—some of us tend to perceive the unbiasedness of an estimator as the most important of the properties, in the sense that if an estimator is biased, then it would be almost pointless to pursue its remaining properties. This problem shows how wrong this attitude may be: be aware that controlling bias does not guarantee that MSE is controlled. In particular, it is sometimes the case that a trade-off occurs between variance and bias in such a way that a small increase in bias can be traded for a larger decrease in variance, resulting in an improvement in MSE. Let's emphasize however that the above example does not imply that $S^2$ should be abandoned as an estimator of $\sigma^2$. Even though on the average, $\hat{\sigma}^2$ will be closer to $\sigma^2$ than $S^2$ if MSE is used as a measure, $\hat{\sigma}^2$ is biased and will, on the average, underestimate $\sigma^2$. This fact alone may make us uncomfortable about using $\hat{\sigma}^2$ as an estimator of a $\sigma^2$. Furthermore, it can be argued that MSE, while a reasonable criterion for location parameters (e.g., the mean), is not reasonable for scale parameters, so the above comparison should not even be made: MSE penalizes equally for overestimation and underestimation, which is fine in the location case. In the scale case (e.g., for the variance), however, 0 is a natural lower bound, so the estimation problem is not symmetric. Use of MSE in this case tends to be forgiving of underestimation.

**10. ($S^2$ is not UMVUE)** Let $X_1$, ..., $X_n$ be a random sample from a $N(\mu, \sigma^2)$ population with both $\mu$ and $\sigma^2$ unknown and consider the sample variance estimator for $\sigma^2$. 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

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

**Solution**. Because we know that $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$, therefore

$$E[S^2] = E\left[\frac{\sigma^2}{n-1}\chi^2_{n-1}\right] = \frac{\sigma^2}{n-1}E[\chi^2_{n-1}] = \sigma^2,$$

which establishes that $S^2$ is unbiased. Given the functional form of a normal PDF/joint density function, then setting $\boldsymbol{\theta} = \sigma^2$,

$$\begin{aligned}
\frac{\partial^2}{\partial(\sigma^2)^2} \ln f(x;\sigma^2) &= \frac{\partial^2}{\partial(\sigma^2)^2} \ln\left\{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}\right\} \\
&= \frac{\partial^2}{\partial(\sigma^2)^2}\left\{-\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln\sigma^2 - \frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\} \\
&= \frac{\partial}{\partial\sigma^2}\left\{-\frac{1}{2\sigma^2} + \frac{1}{2}\frac{(x-\mu)^2}{\sigma^4}\right\} = \frac{1}{2\sigma^4} - \frac{(x-\mu)^2}{\sigma^6}.
\end{aligned}$$

Exploting the hint, we have:

$$\begin{aligned}
-nE\left\{\frac{\partial^2}{\partial(\sigma^2)^2}\ln f(x;\sigma^2)\right\} &= -nE\left[\frac{1}{2\sigma^4} - \frac{(x-\mu)^2}{\sigma^6}\right] \\
&= -\frac{n}{2\sigma^4} + \frac{n\sigma^2}{\sigma^6} = \frac{-n+2n}{2\sigma^4} = \frac{n}{2\sigma^4}.
\end{aligned}$$

At this point, the Cramer-Rao lower bound is:

$$Var[S^2] \geq \frac{\left(\frac{d}{d\sigma^2}E[S^2]\right)^2}{-E\left\{\frac{\partial^2}{\partial(\sigma^2)^2}\ln f(x;\sigma^2)\right\}} = \frac{\left(\frac{d}{d\sigma^2}\sigma^2\right)^2}{\frac{n}{2\sigma^4}} = \frac{2\sigma^4}{n}.$$

However we know from problem 5b that

$$Var[S^2] = \frac{2\sigma^4}{n-1} > \frac{2\sigma^4}{n}$$

which shows that $S^2$ does not reach the bound.

*The lesson of this exercise is:* Do not assume that because you use some estimator very frequently in your every day life, this must be the best available around. In fact, Casella and Berger (2002, p. 341) prove that the best unbiased estimator of $\sigma^2$ is $\tilde{\sigma}^2 \equiv n^{-1}\sum_{i=1}^{n}(x_i-\mu)^2$ which is the ML estimator calculable only if $\mu$ is known. If $\mu$ is unknown, the bound cannot be attained.

**11. (Uniqueness of the Best Unbiased Estimator)** Show that if $W$ is BUE for $\theta$, 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 $Cov[X,Y] \leq \sqrt{Var[X]Var[Y]}$).

**Solution**. Suppose $W^*$ is another BUE, which implies that $Var[W] = Var[W^*]$, and consider a third estimator $W' = 0.5W + 0.5W^*$. Note that

$$\begin{aligned} Var[W'] &= Var\left[\frac{1}{2}W + \frac{1}{2}W^*\right] = \frac{1}{4}Var[W] + \frac{1}{4}Var[W^*] + \frac{1}{2}Cov[W,W^*] \\ &\leq \frac{1}{4}Var[W] + \frac{1}{4}Var[W] + \frac{1}{2}\sqrt{Var[W]Var[W^*]} \text{ (by Cauchy-Schwarz's inequality)} \\ &= \frac{1}{4}Var[W] + \frac{1}{4}Var[W] + \frac{1}{2}Var[W] = Var[W]. \end{aligned}$$

But if $Var[W'] < Var[W]$, we have a contradiction already; therefore it must be $Var[W'] = Var[W]$. If so, then it must be $Var[W] = Var[W^*]$ so that

$$W^* = a(\theta) + W.$$

Yet, $E[W^*] = a(\theta) + E[W] = E[W] = \theta$ which implies that $a(\theta) = 0$. Hence $W^* = W$ and the BUE can only be unique.

**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.

**Solution**. Recall that $\bar{X}_n \sim N(\mu, 1/n)$ so that

$$\begin{aligned} \Pr\left(|\bar{X}_n - \mu| < \epsilon\right) &= \int_{\mu-\epsilon}^{\mu+\epsilon} \frac{n^{1/2}}{(2\pi)^{1/2}} \exp\left[-\frac{n(\bar{x}_n-\mu)^2}{2\sigma^2}\right] d\bar{x}_n \\ &= \int_{-\epsilon}^{\epsilon} \frac{n^{1/2}}{(2\pi)^{1/2}} \exp\left[-\frac{ny^2}{2\sigma^2}\right] dy \quad \text{(define } y \equiv \bar{x}_n - \mu\text{)} \\ &= \int_{-\epsilon\sqrt{n}}^{\epsilon\sqrt{n}} \frac{1}{(2\pi)^{1/2}} \exp\left[-\frac{t^2}{2\sigma^2}\right] dy \quad \text{(define } t \equiv \sqrt{n}y\text{; Jacobian is } n^{-1/2}\text{)} \\ &= \Pr(-\epsilon\sqrt{n} < Z < \epsilon\sqrt{n}) \overset{n\to\infty}{\to} 1. \end{aligned}$$

*References*

Casella, G., and R.L., Berger, *Statistical inference*, Duxbury Press, 2002.