Stat 135 Summer 2017 – Homework 1

Bias, Method of Moments, Standard Error, Confidence intervals, moment generating functions, MLE, Bootstrapping

Only submit starred problem. Others are for extra practice.

Read sections 4.5, 7.1-7.3, 8.1-8.5.1 of the textbook *Mathematical Statistics and Data Analysis, 3rd ed.*by John Rice. To obtain full credit, please **write clearly** and **show your reasoning**.

Rice Chapter 7 problems
2*, 4, 8b*, 10, 11*, 15a, 24 (requires the method of Lagrange multipliers), 36*

Problem 1A (expectation and p.d.f. of a *function* of a random variable, $Y = g(X)$). Consider a random variable $X \sim \text{Uniform}[e, e^2]$ and define the new random variable $Y = \ln X$.

(a) Compute $E(Y)$, using the theorem that states that $E(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx$.

(b) Now we will calculate $E(Y)$ by computing the *probability density function* of Y *first*. To do so the initial step is to figure out what the c.d.f. of Y is, by noticing that:

$$F_Y(y) = P(Y \le y) = P(\ln X \le y) \overset{(*)}{=} P(X \le e^y) = F_X(e^y). \tag{1}$$

Note that in step $(*)$ we have applied the exponential function to both sides of the inequality, and used the fact that such function is *monotone increasing* (that is, $a \le b$ if and only if $e^a \le e^b$. Had we used a monotone decreasing function, we would have had to switch the direction of the inequality!). To find the p.d.f. of Y apply the chain rule[1] to (1). Now that you have f_Y, compute the expectation of Y via the usual formula $E(Y) = \int_{-\infty}^{\infty} y f_Y(y)\, dy$. You should get the same result as in part **(a)**.

(c) Now let's spice things up a little and use a function g that is not monotone increasing or decreasing: consider a standard normal random variable $Z \sim \mathcal{N}(0, 1)$, and[2] $T = g(Z) = Z^2$. First note that $T \ge 0$, so its p.d.f. $f_T(t)$ will be zero for $t < 0$. To compute it, we first find the c.d.f.; for $t \ge 0$ we will have:

$$F_T(t) = P(T \le t) = P(Z^2 \le t) = P(-\sqrt{t} \le Z \le \sqrt{t}) = \Phi(\sqrt{t}) - \Phi(-\sqrt{t}).$$

Finally, use the chain rule to compute the p.d.f. f_T, and roughly <u>sketch</u> its graph. This is the so-called χ^2 ("**chi-squared**") probability density function, which is of fundamental importance in statistics.

Problem 1B* (properties of the moment generating function). We defined the moment generating function (m.g.f.) of a random variable X as $\psi_X(t) = E(e^{Xt})$ which is a function of t (note: I sometimes use $\psi_X(t)$ instead of $M_X(t)$ as the name of a moment generating function). More explicitly, it has one of the two forms:

$$\psi_X(t) = E(e^{Xt}) = \sum_{\text{all}\, x} e^{xt} P(X = x) \qquad \text{or} \qquad \psi_X(t) = E(e^{Xt}) = \int_{-\infty}^{\infty} e^{xt} f_X(x)\, dx,$$

if X is a *discrete* or *continuous* random variable, respectively. (In case you are asking yourself about the convergence of the summation or the integral, one can prove, using the theory of Laplace transforms, that they converge at least in an interval of $\mathbb{R}$ that contains the point $t = 0$; we take this interval as the domain of definition of the moment generating function $\psi_X(t)$).

[1] Remember that the *chain rule* gives the derivative of "nested" functions: $\frac{d}{dx}\big(h(\ell(x))\big) = h'(\ell(x))\, \ell'(x)$.

[2] Note that in this case $g(z) = z^2$ is a *nonlinear* function of z so you should *not* expect the random variable $T = Z^2$ to retain a Normal pdf.

(a) Show that, for any r.v. X, it is the case that $\psi_X(0) = 1$. (So the m.g.f. is always defined at $t = 0$.)

(b) Show that if $Y = aX + b$, where a and b are constants, then $\psi_Y(t) = e^{bt}\psi_X(at)$.

(c) Show that if the random variables X and Y are *independent* and $Z = X+Y$, then $\psi_Z(t) = \psi_X(t)\psi_Y(t)$. *Hint:* you have to use the fact that if two random variables X and Y are independent and f and g are two functions, then the two new random variables $U = f(X)$ and $V = g(Y)$ are also independent.

Problem 1C*. Another property of the moment generating function is that if two random variables have the same m.g.f., then they have the same p.d.f. (we say that the mapping p.d.f. $\mapsto$ m.g.f. is one-to-one). In class we showed that if we have $X \sim \text{Gamma}(\alpha, \lambda)$, then $\psi_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^\alpha$. Use these facts and part **(c)** of the previous problem to show that if $X \sim \text{Gamma}(\alpha, \lambda)$ and $Y \sim \text{Gamma}(\beta, \lambda)$ are independent random variables and we define $Z = X + Y$, then $Z \sim \text{Gamma}(\alpha + \beta, \lambda)$.

Problem 1D* (m.g.f. of Binomial random variables). We proved that the m.g.f. $\psi_X(t)$ "generates" the moments of the random variable X by *differentiation*, and computation at $t = 0$ (rather than by integration or summation, which is typically harder). For example,

$$\psi'_X(0) = E(X), \quad \psi''_X(0) = E(X^2), \quad \ldots, \quad \psi^{(n)}_X(0) = E(X^n),$$

where the superscript $^{(n)}$ indicates the n^{th} derivative.

(a) Assume that $X \sim \text{Binomial}(n, p)$, i.e. with p.f. $f_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$, for $k = 0, 1, \ldots, n$. Find the function $\psi_X(t)$, in terms of the parameters n and p.

Hint: You will need to use the binomial theorem, which states that $\displaystyle\sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} = (a+b)^n$.

(b) Find $E(X)$ and $\text{Var}(X)$ using the method described above.

Note that this is *much* simpler than computing, for example, $E(X) = \displaystyle\sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k}$.

Problem 1E (m.g.f. of Normal random variables).

(a) Consider a *standard* Normal random variable $Z \sim \mathcal{N}(0, 1)$, and compute its m.g.f. ψ_Z. *Hint:* You need to "complete the square" at the exponent: $\frac{z^2}{2} - tz = \frac{1}{2}(z^2 - 2tz + t^2) - \frac{t^2}{2} = \frac{1}{2}(z - t)^2 - \frac{t^2}{2}$, and then perform a change of variable in the integral.

(b) Now use part (b) of Problem 1B to compute the m.g.f. of a *generic* Normal r.v. $X \sim N(\mu, \sigma^2)$.

(c) Finally, use part (c) of Problem 1B to show that if $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ are independent and we define $Z = X + Y$, then $Z \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Remark: it should be clear by now that using the moment generating functions to prove the above fact is a much simpler method than computing the convolution between two Normal p.d.f.'s!

Problem 1F. Suppose that X is a random variable for which the m.g.f. is as follows:

$$\psi_X(t) = \frac{1}{6}\left(4 + e^t + e^{-2t}\right).$$

Find the probability distribution of X. *Hint:* It is a simple discrete distribution for example, pmf $P(X = 0) = .3$, $P(X = 1) = .7$.

Problem 1G* (chi-square random variables with n degrees of freedom). Given a standard Normal r.v. $Z \sim N(0,1)$, let us define the new random variable $T = Z^2$. In Problem 1A you computed the p.d.f. of such random variable, called a χ^2 (**chi-square**) p.d.f., namely

$$f_T(t) = \begin{cases} \dfrac{1}{\sqrt{2\pi}} \dfrac{1}{\sqrt{t}} e^{-\frac{t}{2}} & \text{for } t \geq 0, \\ \\ 0 & \text{otherwise.} \end{cases}$$

If we compare this with the general form of an Gamma(α, λ) p.d.f., we recognize that the χ^2 probability density function is, in fact, a Gamma$\left(\frac{1}{2}, \frac{1}{2}\right)$ density.

(a) In light of this, if $T \sim \chi^2$, what are $E(T)$ and $\mathrm{Var}(T)$?

(b) If $Z_1, Z_2, \ldots, Z_n$, are i.i.d. standard normal random variables, let us define $T_n = Z_1^2 + Z_2^2 + \ldots + Z_n^2$. We say that it is a χ^2_n random variable (**chi-square with n degrees of freedom**). What is the expression of the probability density function of a χ^2_n random variable? (You should be able to answer this question without performing any significant computation!). What is its m.g.f.?

(c) What are $E(T_n)$ and $\mathrm{Var}(T_n)$?

(d) For large values of n, what is a good approximate probability model for T_n (Hint: apply central limit theorem)?

Problem 1H* (intro to MLE). Suppose that X is a discrete random variable with probability function

$$P(X = 0 \,|\, \theta) = \frac{2}{3}\theta, \qquad P(X = 1 \,|\, \theta) = \frac{1}{3}\theta, \qquad P(X = 2 \,|\, \theta) = \frac{2}{3}(1-\theta), \qquad P(X = 3 \,|\, \theta) = \frac{1}{3}(1-\theta),$$

where $0 \leq \theta \leq 1$ is an unknown parameter. The following 10 independent observations were taken from such a distribution: (3, 0, 2, 1, 3, 2, 1, 0, 2, 1). (Equivalently, these are the values of a sample of ten i.i.d. random variables $X_1, \ldots, X_{10}$, each with the above probability function.)

(a) What is the likelihood function $\ell(\theta)$ for the above sample? (Note: in class I called this $lik(\theta)$.) What is the log-likelihood function $L(\theta)$? (Note: in class I called this $l(\theta)$).

(b) Compute the value of θ that maximizes the probability of the above sample: that is, compute the maximum likelihood estimate $\widehat{\theta}_{\mathrm{ML}}$ for the above sample (by setting $L'(\theta) = 0$, and then solving for θ).

(c) Now suppose that we have a *generic* sample of size n, i.e. of the type $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$, where $x_i \in \{0, 1, 2, 3\}$. Express the likelihood function $\ell(\theta)$ and the log-likelihood $L(\theta)$ in terms of

$$\begin{aligned} n_0 &= \text{\# of samples that are equal to 0,} \\ n_1 &= \text{\# of samples that are equal to 1,} \\ n_2 &= \text{\# of samples that are equal to 2,} \\ n_3 &= \text{\# of samples that are equal to 3.} \end{aligned}$$

Note that $n = n_0 + n_1 + n_2 + n_3$ (the total number of samples). Find $\widehat{\theta}_{\mathrm{ML}}$ in terms of the above n_i's.

Problem 1I* (MLE for the Poisson distribution). Consider n i.i.d. random variables, each of them Poisson(λ), with *unknown* parameter λ. Suppose that we measure the values $X_1 = x_1$, $X_2 = x_2$, $\ldots$, $X_n = x_n$. The likelihood function is $\ell(\lambda) = f(k_1|\lambda) \cdot \ldots \cdot f(k_n|\lambda)$, where $f(k|\lambda) = e^{-\lambda}\frac{\lambda^k}{k!}$, with $k = 0, 1, 2, 3, \ldots$. Find the value of λ that maximizes the probability of the data, i.e. the maximum likelihood estimate $\widehat{\lambda}_{\mathrm{ML}}$. This is done by first computing the log-likelihood $L(\lambda) = \log \ell(\lambda)$, by setting $L'(\lambda) = 0$ and solving for λ. Why is the result not that surprising? (Remember that if $X_i \sim$ Poisson(λ), then $E(X_i) = \lambda$.)

Problem 1J* (MLE for the Uniform p.d.f.). Consider n i.i.d. continuous random variables, each of them Uniform$(0, \theta)$, with *unknown* parameter θ. Suppose that we measure the values $X_1 = x_1$, $X_2 = x_2$, ..., $X_n = x_n$ (here, $x_i > 0$ for all i). Compute the likelihood function for such data sample, and compute $\widehat{\theta}_{\text{ML}}$. *Hint:* It is a bit tricky—in this case, the maximum likelihood estimate is found *without* differentiation.

Problem 1K Consider n i.i.d. random variables $X_1, X_2, \ldots, X_n \sim$ Geometric(p), where $p \in [0, 1]$ is an unknown parameter. Find the Maximum Likelihood Estimate (MLE) of p.

Problem 1L* Let's consider Problem 1H in this homework assignment. We had n i.i.d. random variables X_1, X_2, ..., X_n, each with probability function

$$P(X = 0 \,|\, \theta) = \frac{2}{3}\theta, \quad P(X = 1 \,|\, \theta) = \frac{1}{3}\theta, \quad P(X = 2 \,|\, \theta) = \frac{2}{3}(1 - \theta), \quad P(X = 3 \,|\, \theta) = \frac{1}{3}(1 - \theta), \tag{2}$$

where θ is an unknown parameter. Assuming that n_i = "# of samples that are equal to i", for $i = 0, 1, 2, 3$ (note that $n_0 + n_1 + n_2 + n_3 = n$) you found that the Maximum Likelihood Estimate (MLE) of θ is given by $\widehat{\theta}_{\text{ML}} = \frac{n_0+n_1}{n}$. If we want to study its statistical properties then we must write it as a *random variable*, i.e.

$$\widehat{\theta}_{\text{ML}} = \frac{N_0 + N_1}{n} \tag{3}$$

where we have used the random variables N_i = "# of observations in the sample that are equal to i", whose probability distributions depend on (4).

(a) Show that (3) is an *unbiased* estimator of θ, i.e. that $E(\widehat{\theta}_{\text{ML}}) = \theta$. *Hint:* First find the probability function of the random variable $Y = N_0 + N_1$, which is the number of elements of the sample X_1, X_2, ..., X_n that are equal to *either* 0 *or* 1 (note that n is fixed). What is $P(Y = k)$, for $k = 0, 1, \ldots, n$?

(b) Compute $\text{Var}(\widehat{\theta}_{\text{ML}})$ in terms of the parameter θ.

Problem 1M* (bootstapping the SE of an MLE estimate) Let's consider Problem 1H in this homework assignment. We had n i.i.d. random variables X_1, X_2, ..., X_n, each with probability function

$$P(X = 0 \,|\, \theta) = \frac{2}{3}\theta, \quad P(X = 1 \,|\, \theta) = \frac{1}{3}\theta, \quad P(X = 2 \,|\, \theta) = \frac{2}{3}(1 - \theta), \quad P(X = 3 \,|\, \theta) = \frac{1}{3}(1 - \theta), \tag{4}$$

where θ is an unknown parameter.
Use R to bootstrap an approximate standard error of the maximum likelihood estimate.

Problem 1N*. Given n i.i.d. random variables $X_1, X_2, \ldots, X_n$, each with mean μ and variance $\sigma^2 > 0$ (both unknown parameters), we saw in class that the *sample mean* $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is an unbiased estimator of the mean μ, whereas the so-called *sample variance*

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

is an unbiased estimator of the variance σ^2. In other words, $E(s^2) = \sigma^2$. Is $s = \sqrt{s^2}$ an *unbiased* estimator of the standard deviation σ? If not, does s tend to underestimate or overestimate the standard deviation? *Hint:* Very little computation is needed. Start by writing the formula for computing the variance of s.

Problem 1O. Suppose that $X_1, \ldots, X_n$ form a random i.i.d. sample from the continuous uniform p.d.f. on the interval $[0, \theta]$, where the value of the parameter θ is unknown. Let $Y_n = \max(X_1, \ldots, X_n)$; as you found in Problem 1J of the previous homework assignment, Y_n is in fact the Maximum Likelihood Estimator of θ.

(a) Show that $E(Y_n) = \dfrac{n}{n+1}\,\theta$, i.e. Y_n is a *biased* estimate of θ.

(Note, however, that it becomes essentially unbiased for a large sample size, since $\frac{n}{n+1} \simeq 1$ as $n \to \infty$.) *Hint:* First find the c.d.f. of Y_n, i.e. the function $F_{Y_n}(y) = P(Y_n \leq y)$, which obviously depends on θ. Then compute the p.d.f. $f_{Y_n}(y)$ by differentiation with respect to y. Finally, compute the expected value $E(Y_n) = \int_{-\infty}^{\infty} y\, f_{Y_n}(y)\, dy$.

(b) Find a multiplicative constant c such that $Z_n = cY_n$ is an unbiased estimate of θ (this is much simpler).

The Method of Moments (MM), in the general case. If we want to estimate the ℓ parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_\ell)$ of a probability distribution $f(x|\theta_1, \theta_2, \ldots, \theta_\ell)$ from independent samples $X_1, X_2, \ldots, X_n$ drawn precisely from such distribution, we follow three steps:
(1) We compute the expressions for the first ℓ moments, $\mu_\ell = E(X^\ell)$, $k = 1, \ldots, \ell$:

$$\begin{cases} \mu_1 &= g_1(\theta_1, \theta_2, \ldots, \theta_\ell) \\ \vdots &= \vdots \\ \mu_\ell &= g_\ell(\theta_1, \theta_2, \ldots, \theta_\ell) \end{cases}$$

This operation requires either integration, or summation, or the use of the moment generation function. For example, if $X \sim \text{Gamma}(\alpha, \lambda)$ we have $\ell = 2$, $\mu_1 = \frac{\alpha}{\lambda}$ and $\mu_2 = \frac{\alpha(\alpha+1)}{\lambda^2}$.
(2) Using (typically pretty simple) algebra, we *invert* the above system of equations, and express the parameters $\theta_1, \theta_2, \ldots, \theta_\ell$ in terms of the moments $\mu_1, \mu_2, \ldots, \mu_\ell$:

$$\begin{cases} \theta_1 &= h_1(\mu_1, \mu_2, \ldots, \mu_\ell) \\ \vdots &= \vdots \\ \theta_\ell &= h_\ell(\mu_1, \mu_2, \ldots, \mu_\ell) \end{cases} \tag{5}$$

For example, if $X \sim \text{Gamma}(\alpha, \lambda)$, it is the case that $\alpha = \frac{\mu_1^2}{\mu_2 - \mu_1^2}$ and $\lambda = \frac{\mu_1}{\mu_2 - \mu_1^2}$.
(3) Finally, we insert into the functions (5) the following estimators of the moments:

$$\widehat{\mu}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \widehat{\mu}_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2, \quad \ldots \quad \widehat{\mu}_\ell = \frac{1}{n}\sum_{i=1}^{n} X_i^\ell,$$

(note: $\widehat{\mu}_1 = \overline{X}$, i.e. it is the *sample mean*) to define the MM estimators for the parameters:

$$\begin{cases} \widehat{\theta}_1 &= h_1(\widehat{\mu}_1, \widehat{\mu}_2, \ldots, \widehat{\mu}_\ell) \\ \vdots &= \vdots \\ \widehat{\theta}_\ell &= h_\ell(\widehat{\mu}_1, \widehat{\mu}_2, \ldots, \widehat{\mu}_\ell) \end{cases}$$

In the Gamma(α, λ) case, we get: $\widehat{\alpha} = \frac{(\widehat{\mu}_1)^2}{\widehat{\mu}_2 - (\widehat{\mu}_1)^2} = \frac{\overline{X}^2}{\widehat{\mu}_2 - \overline{X}^2}$ and $\widehat{\lambda} = \frac{\widehat{\mu}_1}{\widehat{\mu}_2 - (\widehat{\mu}_1)^2} = \frac{\overline{X}}{\widehat{\mu}_2 - \overline{X}^2}$.

Problem 1P*. Consider the i.i.d. random variables $X_1, X_2, \ldots, X_n$. Show that $\widehat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ is an unbiased estimator of the k^{th} moment $\mu_k = E(X^k)$; i.e. that $E(\widehat{\mu}_k) = \mu_k$, for any natural number k.
Remark: Despite this the parameter estimators that are obtained through the method of moments are sometimes biased (that happens because the functions (5) are often nonlinear).

Problem 1Q. Consider the i.i.d. random variables $X_1, X_2, \ldots, X_n \sim \text{Geometric}(p)$, with p unknown. Find $\widehat{p}_{\text{MM}}$, i.e. the method of moments estimator of p (in this case $\ell = 1$, so you need only one moment).

Problem 1R*. Consider the i.i.d. random variables $X_1, X_2, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$, with both μ and σ^2 unknown. Find $\widehat{\mu}_{\text{MM}}$ and $\widehat{\sigma}^2_{\text{MM}}$ i.e. the method of moments estimators of the mean and the variance, and show that in this case the MM estimators are, in fact, the *same* as the ML estimators (example B page 269 in book). *Remark:* As we know from class, note that the resulting estimator for the variance is *biased*.