Homework Assignment #3 for Prob-Stats, Fall 2018
Due date: Monday, November 19, 2018

Topics: parametric families of distributions; distributions of transformations; hazard rates and optimal waiting times; estimation for parametric and non-parametric decision problems.

Readings: CB: Ch. 2, 3, and 7.2-3; Bierens: Ch. 4, 8.1-3.

A. Casella and Berger, p. 80, on moment generating functions, 2.30, 2.31, and 2.32.

B. For any $r, s \in \mathbb{R}$, let $(r \vee s) = \min(r, s)$ and $(r \wedge s) = \max(r, s)$. It is immediate that $r + s = (r \wedge s) + (r \vee s)$. Let $X, Y : \Omega \to \mathbb{R}$ be two elements of L_1, and show that

$$E\,(X \vee Y) = E\,X + E\,Y - E\,(X \wedge Y).$$

Letting $X = 1_A$ and $Y = 1_B$, what does the previous yield?

C. Let $X : \Omega \to \mathbb{R}$ be a random variable with a continuous cdf, $F(\cdot)$.
   1. Give the cdf of $Y(\omega) := F(X(\omega))$.
   2. Suppose that X_n, $n = 0, 1, 2, \ldots$ are i.i.d. with continuous cdf $F(\cdot)$. Define $T(\omega) = \min\{t \geq 1 : X_t(\omega) > X_0(\omega)\}$. Give $E\,(T|X = x)$.
   3. Give $E\,\tau$.

D. Casella and Berger p. 78 on the most commonly used loss function estimators, 2.18 and 2.19. [This is background for the next problem.]

E. For $x, m \in \mathbb{R}$ and $0 < \alpha < 1$, define

$$\ell_\alpha(x, m) = \begin{cases} \alpha(m - x) & \text{if } x < m, \text{ and} \\ (1 - \alpha)(x - m) & \text{if } x \geq m. \end{cases}$$

For each of the following distributions for X (the notation for the following distributions is from the Casella and Berger Table of Common Distributions, p. 621-626), solve for

$$\widehat{m}_\alpha := \arg\min_m E\,\ell(X, m).$$

   1. $X \sim \text{Poisson}\,(\lambda)$.
   2. $X \sim \text{logistic}\,(\mu, \beta)$.
   3. $X \sim \text{Uniform}\,(a, b)$.
   4. $X \sim \text{Weibull}\,(\gamma, \beta)$.
   5. $X \sim N(\mu, \sigma^2)$.

F. Casella and Berger on hazard rates, 3.25 and 3.26, p. 131-2. [This is background for the next three problems.]

G. People placing phone calls can be of two types, chatty and taciturn. If they are chatty, the random time until they are done talking is distributed as an exponential(μ), if taciturn, as an exponential(λ, where $\lambda < \mu$ so that, on average, the chatty people talk longer, average time $1/\mu$, than the taciturn, average time $1/\lambda$.

1. If a chatty and a taciturn person start talking at the same time, what is the probability that the taciturn person will be done first?
2. If a chatty person has been talking for t minutes before the taciturn person starts talking, what is the probability that the taciturn person will be done first? [This is not an intuitive result, it comes from the memorylessness of the exponential distribution.]
3. If we know that there is a probability α that the person talking is chatty and $(1-\alpha)$ that they are taciturn, and T is the time until they are done, give $P(T > t)$.
4. If $W : \Omega \to [0, \infty)$ has density $f(\cdot)$, then the hazard rate for W is defined as $h(t) = \frac{f(t)}{1-F(t)}$. Give the hazard rate for waiting for a chatty person to be done talking.
5. If we know that there is a probability α that the person talking is chatty and $(1-\alpha)$ that they are taciturn, and T is the time until they are done, give the hazard rate for T. Find $\lim_{t\to\infty} h(t)$ and explain why you get this answer.

H. Suppose that $X \sim$ negative exponential(λ), i.e. $F_X(t) = 1-e^{-\lambda t}$ for some $\lambda > 0$, and that $Y(\omega) := X^{\gamma}(\omega)$ for $\gamma > 0$. [Y belongs to the class of Weibull distributions.]
1. Derive the density and the cdf of Y.
2. Give the hazard rate for Y.
3. When is the hazard rate increasing? Decreasing? Explain why this is true.
4. Give the conditions on β and γ under which $E\left(Y|Y > t\right) - t$ is increasing, and under which it is decreasing. Explain.
5. The cumulative hazard is defined as $H(t) = \int_0^t h(x)\,dx$. Show that $1 - F(t) = e^{-H(t)}$ for any non-negative random variable having a density. Rederive the cdf of Y from this.
6. If $Y : \Omega \to [0, \infty]$ has a density on $[0, \infty)$ and $P(Y = \infty) = q > 0$, show that $P(W = \infty|W > t) \uparrow 1$ as $t \uparrow \infty$. Does this imply that $\lim_{t\uparrow\infty} h(t) = 0$? If not, give (or sketch) an example. If so, give a proof. [Waiting time distributions with $P(Y = \infty) > 0$ are called "incomplete" distributions.]

I. Suppose that we have a list of everyone who in unemployed at a randomly picked time T, which is independent of everything else in sight. We record Y_i, the length of time between becoming unemployed and T, we record X_i, the time between T and their new job, and we construct $T_i = Y_i + X_i$, the total length of i's unemployment spell. Suppose also that the random time between jobs has a negative exponential distribution. Consider three estimators of λ, $\widehat{\theta}_X = \frac{1}{n}\sum_i X_i$, $\widehat{\theta}_Y = \frac{1}{n}\sum_i Y_i$, and $\widehat{\theta}_T = \frac{1}{n}\sum_i T_i$.
1. Find $E_\lambda \widehat{\theta}_X$.
2. Find $E_\lambda \widehat{\theta}_Y$.
3. Find $E_\lambda \widehat{\theta}_T$.

4. Suppose now that the waiting time has a Weibull distribution as specified in the previous problem. If $\gamma > 1$, what pattern would we expect to see in the previous estimators? What about if $\gamma < 1$?

J. At a flow cost of $c \geq 0$, one can keep searching for a source of higher profits (a low cost source of a crucial input, a process breakthrough, a new product). If found, expected net flow profits of $\overline{\pi}$ result. If one abandons the search, the decision is, by assumption, irreversible, and the known alternative yields expected net flow profits of $\underline{\pi}$, $\overline{\pi} > \underline{\pi} > 0$. The random time until which the source is successful is a random variable $Y : \Omega \to [0, \infty]$ that has a density on $[0, \infty)$ and for which it is possible that $P(Y = \infty) = q > 0$. Because we search the most likely places first, we assume that the hazard rate for Y is decreasing, and for convenience, we assume that it is continuous. You choose a time t_1 at which you will abandon searching and accept the lower $\underline{\pi}$. You do this to maximize your expected payoffs where,
   - if $Y > t_1$, then your payoffs are $\int_0^{t_1} -ce^{-rx}\,dx + \underline{\pi}e^{-rt_1}$, and
   - if $Y < t_1$, then your payoffs are $\int_0^{Y} -ce^{-rx}\,dx + e^{-rY}\overline{\pi}$.

   1. Show that the FOCs for your problem have a unique solution at the t_1^* that solves $h(t_1^*) = r(c + \underline{\pi})/(\overline{\pi} - \underline{\pi})$.
   2. Show that the SOCs are satisfied at your solution.
   3. Show that the optimal t_1^* is higher for higher $\overline{\pi}$ and explain the economics of your answer.
   4. Show that the optimal t_1^* is lower for higher values of c and explain the economics of your answer.
   5. Show that the optimal t_1^* is lower for higher values of $\underline{\pi}$ and explain the economics of your answer.

K. Casella and Berger on exponential families and natural parameter spaces, 3.28 and 3.29, p. 132. [The exponential families of distributions are a flexible class, and their maximum likelihood estimators (MLEs) are often easy to calculate. The next problem asks for MLEs for three classes of distributions.]

L. This problem is background for examining the expected utility consequences of model mis-specification. The data, $(X_1, \ldots, X_n)$, is an i.i.d. collection of random integers. We will consider three parametrized classes of distributions for the data: the discrete uniforms, $\{U_N : N \in \mathbb{N}\}$ defined by $U_N(n) = 1/N$ for $n = 1, \ldots, N$; the geometric distributions, $\{G_p : p \in [0, 1]\}$, defined by $G_p(n) = p(1-p)^{n-1}$, $n = 1, 2, \ldots$; and the "Poisson plus 1" distributions, $\{P_\lambda : 0 \leq \lambda < \infty\}$, defined by $P_\lambda(n) = \frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!}$, $x = 1, 2, \ldots$.
   1. Give the likelihood function, the log likelihood function, and the maximum likelihood estimator, $\widehat{N}_n$, associated with the assumption that the data, $(X_1, \ldots, X_n)$, is distributed according to some discrete uniform distribution with unknown N. Calculate the bias of $\widehat{N}_n$ and show that, with probability 1, $\widehat{N}_n$ converges to the true N.

2. Give the likelihood function, the log likelihood function, and the maximum likelihood estimator, $\widehat{p}$, associated with the assumption that the data, $(X_1, \ldots, X_n)$, is distributed according to some geometric distribution with unknown p. Calculate the bias of the estimator and show that, with probability 1, $\widehat{p}_n$ converges to the true p.
3. Give the likelihood function, the log likelihood function, and the maximum likelihood estimator, $\widehat{\lambda}$, associated with the assumption that the data, $(X_1, \ldots, X_n)$, is distributed according to some Poisson distribution with unknown λ. Calculate the bias of the estimator and show that, with probability 1, $\widehat{\lambda}_n$ converges to the true λ.

M. [Problem L cont.] Consider the problem

$$\max_{a\in\{0,1\}} E\, u(a, X) \text{ where } u(a, X) = (1 - a)r \cdot 1_{\{X<7\}} + as \cdot 1_{\{X\geq 7\}}.$$

1. Assuming that X has a discrete uniform distribution with some parameter N, give the value function $V(N) = \max_{a\in\{0,1\}} E\, u(a, X)$ and the optimal strategy for each N.
2. Assuming that X has a geometric distribution with some parameter p, give the value function $V(p) = \max_{a\in\{0,1\}} E\, u(a, X)$ and the optimal strategy for each p.
3. Assuming that X has a exponential distribution with some parameter λ, give the value function $V(\lambda) = \max_{a\in\{0,1\}} E\, u(a, X)$ and the optimal strategy for each λ.

N. [Problem M cont.] Suppose that the data is distributed according to a discrete uniform distribution with unknown parameter N. Suppose also that the decision maker uses the estimator $\widehat{N}_n$ to solve their maximization problem, that is, suppose that after seeing the data, $(X_1, \ldots, X_n)$, they solve

$$V_{n+1} := \max_{a\in\{0,1\}} E\, u(a, X_{n+1})$$

under the assumption that X_{n+1} is distributed according to $U_{\widehat{N}_n}$. Give the distribution of V_{n+1}.

O. [Problem N cont.] Suppose that the decision maker's model is $\{U_N : N \in \mathbb{N}\}$, but that the true distribution is a geometric with parameter p for some $p \in (0, 1)$.
1. Give the distribution of $\widehat{N}_n$ as a function of p.
2. Give $\lim_n \widehat{N}_n$ as a function of p.
3. Suppose that the decision maker uses the estimator $\widehat{N}_n$ to solve their maximization problem, that is, suppose that after seeing the data, $(X_1, \ldots, X_n)$, they solve

$$V_{n+1} := \max_{a\in\{0,1\}} E\, u(a, X_{n+1})$$

under the assumption that X_{n+1} is distributed according to $U_{\widehat{N}_n}$. Give $\lim_n V_{n+1}$.

4. Sketch how your answer would change if the true distribution is a Poisson(λ) for some $\lambda > 0$.

P. [Problem O cont.] Suppose that the decision maker uses the following "empirical strategy," estimate $\widehat{F}_n$ defined by $\widehat{F}_n(m) = \frac{1}{n}\#\{i : X_i = m\}$, and then solves

$$V_{n+1}^{emp} := \max_{a \in \{0,1\}} E\, u(a, X_{n+1})$$

under the assumption that X_{n+1} is distributed according to the cdf $\widehat{F}_n$. Show that, provided that the data is i.i.d. with distribution Q, $V_{n+1}^{emp} \to V(Q)$ where $V(Q) = \max_{a \in \{0,1\}} \sum_x u(a, x) Q(x)$.