Statistics 581, Problem Set 6 Solutions

Wellner; 11/07/2018

1. A. Compute and plot the score for location $-f'(x)/f(x)$ when:
   (a) $f = \phi$, the standard normal density;
   (b) $f(x) = \exp(-x)/(1+\exp(-x))^2$ (logistic);
   (c) $f(x) = (1/2)\exp(-|x|)$ (double exponential);
   (d) $f(x) = t_k$, the $t-$density with $k-$degrees of freedom;
   (e) $f(x) = \exp(-x)\exp(-\exp(-x))$;
   (f) $f(x) = 2\phi(x)\Phi(ax)$ where $\Phi(x)$ is the standard normal d.f. and $a > 0$;
   (g) $f(x) = 1/(\pi(1+x^2))$, the standard Cauchy density.

   B. A density f is called *log-concave* if $\log f$ is a concave function. Let $s < 0$. A density f is called $s-$*concave* if f^s is convex.
   Which of the densities in (a) - (f) are log-concave?
   Which of the densities in (a) - (f) are $s-$concave for some $s < 0$?
   Which of the densities in (a) - (f) are symmetric about 0?

   Solution: A.
   (a) For the normal density $f = \phi$, $-\log f(x) = (1/2)\log(2\pi) + x^2/2$, so $(-\log f)'(x) = x$, and $(-\log f)''(x) = 1$.
   (b) For the logistic density $f(x) = e^{-x}/(1+e^{-x})^2$, so

$$\begin{aligned}
(-\log f(x)) &= x + 2\log(1+e^{-x}),\\
(-\log f(x))' &= 1 - 2e^{-x}/(1+e^{-x}) = (1-e^{-x})/(1+e^{-x})\\
&= F(x) - (1-F(x)) = 2F(x) - 1,\\
(-\log f(x))'' &= 2f(x) \ge 0.
\end{aligned}$$

   (c) For the Laplace density $f(x) = (1/2)\exp(-|x|)$, so

$$\begin{aligned}
(-\log f(x)) &= \log(2) + |x|,\\
(-\log f(x))'(x) &= \mathrm{sign}(x)1\{x \ne 0\} + \mathrm{undefined}1\{x = 0\},\\
(-\log f(x))''(x) &= 0 \cdot 1\{x \ne 0\} + \mathrm{undefined}1\{x = 0\}.
\end{aligned}$$

   (d) For the t_k- distribution $f_k(x) = c_k(1+x^2/k)^{-(k+1)/2}$, so

$$\begin{aligned}
(-\log f(x)) &= -\log c_k + \frac{k+1}{2}\log(1+x^2/k),\\
(-\log f(x))' &= \frac{(k+1}{2}\frac{2x/k}{1+x^2/k} = \frac{k+1}{k}\frac{x}{(1+x^2/k)},\\
(-\log f(x))'' &= = \frac{k+1}{k}\left\{\frac{1-x^2/k}{(1+x^2/k)^2}\right\}
\end{aligned}$$

   (e) For the Gumbel density $f(x) = \exp(-x - e^{-x})$,

$$\begin{aligned}
(-\log f(x)) &= x + e^{-x},\\
(-\log f(x))' &= 1 - e^{-x},\\
(-\log f(x))'' &= e^{-x}.
\end{aligned}$$

(f) For the skew-normal density $f_a(x) = 2\phi(x)\Phi(ax)$, so that

$$\begin{aligned}(-\log f(x)) &= -\log 2 - \log\phi(x) - \log\Phi(ax),\\ (-\log f(x))' &= x - \frac{a\phi(ax)}{\Phi(ax)},\\ (-\log f(x))'' &= 1 + \frac{a^2 x\phi(x)}{\Phi(ax)} + \frac{a^2\phi^2(ax)}{\Phi(ax)^2}.\end{aligned}$$

(g) The Cauchy density $f(x) = 1/(\pi(1+x^2))$ is a special case of the t_k density in (d): thus the score for location is $-(f'/f)(x) = x/(1+x^2)$. Note that this is not a monotone function of x, and hence the Cauchy density is not log-concave. It is $s-$concave with $s = -1/2$.

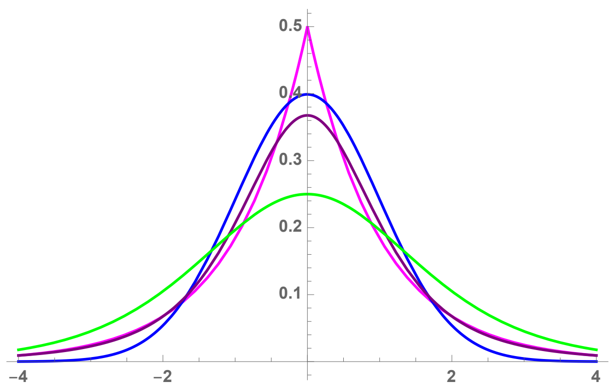

Figure 1: Symmetric densities: (a)-blue, Gaussian; (b)-green, logistic; (c)-magenta, Laplace; (d)-purple, Student t_3

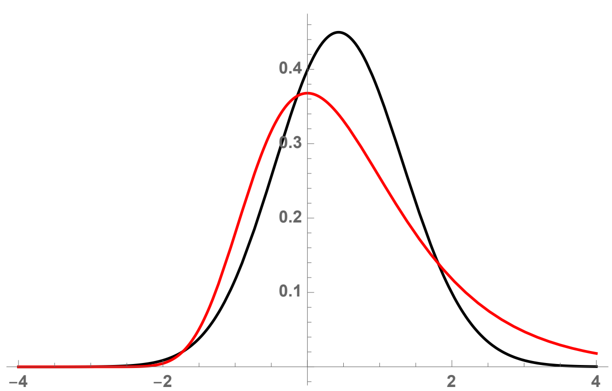

Figure 2: Asymmetric densities: (e)-red, Gumbel; (f)-black, skew-normal, $a = .7$

B. The normal, logistic, double exponential (or Laplace), Gumbel, and skew-normal densities in (a), (b), (c), (e), and (f) are all log-concave with monotone increasing score functions for location. The t_k density in (d) is *not* log-concave. But the t_k density is $s-$concave with $s = -1/(k+1)$ since with $f_k(x) = c_k(1+x^2/k)^{-(k+1)/2}$ we have

$$f_k(x)^s = c_k^{-1/(k+1)} \cdot (1+x^2/k)^{1/2},$$

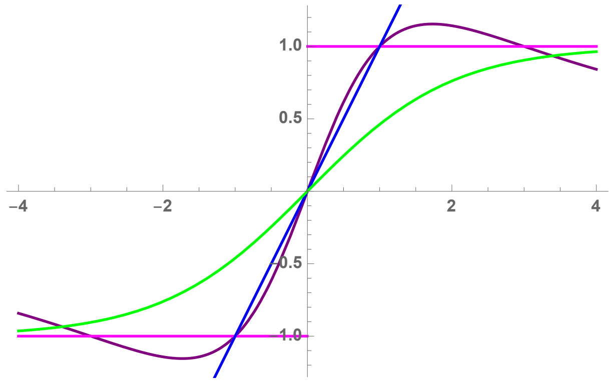

Figure 3: Score functions for location, symmetric densities: (a)-blue, Gaussian; (b)-green, logistic; (c)-magenta, Laplace; (d)-purple, Student t_3

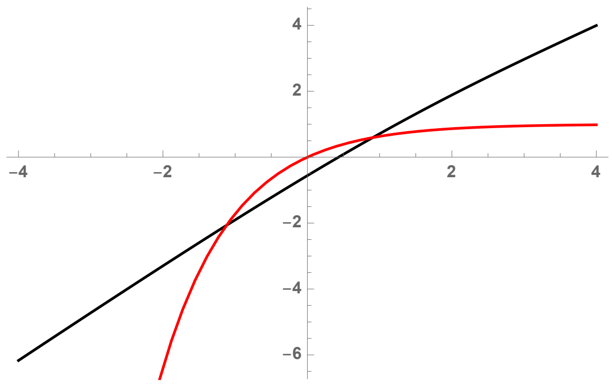

Figure 4: Score functions for location, asymmetric densities: (e)-red, Gumbel; (f)-black, skew-normal, $a = .7$

which is a convex function (of x). The Cauchy density is the special case of the t_k family with $k = 1$, and hence is $s-$concave with $s = -1/2$.
The normal, logistic, Laplace, and t_k densities are all symmetric about 0, while the Gumbel and skew-normal densities are not symmetric.

2. Suppose that $Z \sim N(0,1)$ and, for $\mu \in R$ and $\sigma > 0$, that $X = \mu + \sigma Z \sim P_{\mu,\sigma} = N(\mu, \sigma^2)$.
(a) Compute the likelihood ratio

$$\frac{dP_{\mu,\sigma}}{dP_{0,\sigma}}(x) = \frac{\sigma^{-1}\phi((x-\mu)/\sigma)}{\sigma^{-1}\phi(x/\sigma)} \qquad \text{and} \qquad Y \equiv \log \frac{dP_{\mu,\sigma}}{dP_{0,\sigma}}(X)\,.$$

What is the distribution of Y under $P_{0,\sigma}$ and under $P_{\mu,\sigma}$?
(b) Plot the function

$$l(\mu; X) \equiv \log \frac{dP_{\mu,\sigma}}{dP_{0,\sigma}}(X)$$

as a function of μ.
(c) Find the maximum value of the function $l(\mu; X)$ in (b) (as a function of μ) and the value of $\mu \equiv \hat{\mu}$ which achieves the maximum.
(d) What is the distribution of $\hat{\mu}$ under $P_{0,\sigma}$ and under $P_{\mu,\sigma}$? What is the distribution of $l(\hat{\mu}; X)$ under $P_{0,\sigma}$ and under $P_{\mu,\sigma}$?

Solution: (a) The likelihood ratio

$$\begin{aligned}\frac{dP_{\mu,\sigma}}{dP_{0,\sigma}}(x) &= \frac{\sigma^{-1}\phi((x-\mu)/\sigma)}{\sigma^{-1}\phi(x/\sigma)} = \frac{\exp(-(x-\mu)^2/(2\sigma^2))}{\exp(-x^2/(2\sigma^2))}\\ &= \exp\left(\frac{\mu}{\sigma^2}x - \frac{1}{2}\frac{\mu^2}{\sigma^2}\right).\end{aligned}$$

Hence

$$Y \equiv \log\frac{dP_{\mu,\sigma}}{dP_{0,\sigma}}(X) = \frac{\mu}{\sigma}\frac{X}{\sigma} - \frac{1}{2}\frac{\mu^2}{\sigma^2}.$$

Under $P_{0,\sigma}$ we find that $E(Y) = 0 - \frac{\mu^2}{2\sigma^2}$ and $Var(Y) = \mu^2/\sigma^2 \equiv V^2$ so that

$$Y \sim N(-\frac{1}{2}V^2, V^2) \qquad \text{under } \ P_{0,\sigma}.$$

Under $P_{\mu,\sigma}$ a similar computation gives $E(Y) = \mu^2/\sigma^2 - \mu^2/(2\sigma^2) = V^2/2$ and $Var(Y) = V^2$, so

$$Y \sim N(\frac{1}{2}V^2, V^2) \qquad \text{under } \ P_{\mu,\sigma}.$$

(b) and (c). The function

$$l(\mu,\sigma;X) \equiv \log\frac{dP_{\mu,\sigma}}{dP_{0,\sigma}}(X) = \frac{\mu}{\sigma}\frac{X}{\sigma} - \frac{\mu^2}{2\sigma^2} = \frac{X^2}{2\sigma^2} - \frac{1}{2}\frac{(X-\mu)^2}{\sigma^2}$$

is quadratic in μ with maximum value $X^2/(2\sigma^2)$ which is achieved at $\mu = \hat{\mu} \equiv X$. D. Under $P_{0,\sigma}$, $\hat{\mu} = X \sim N(0,\sigma^2)$ and $l(\hat{\mu},\sigma;X) = X^2/(2\sigma^2) \sim \chi_1^2/2$. Under $P_{\mu,\sigma}$, $\hat{\mu} = X \sim N(\mu,\sigma^2)$ and $l(\hat{\mu},\sigma;X) = X^2/(2\sigma^2) \sim \chi_1^2(\delta)/2$ with $\delta = \mu^2/\sigma^2$.

3. Suppose that $X, X_1, X_2, \ldots, X_n$ are independent Exponential(λ) random variables:

$$P(X \geq x) = \exp(-\lambda x), \qquad x > 0.$$

(a) Show that the r-th moment of X, $\mu_r \equiv \mu_r(\lambda)$ is given by

$$\mu_r(\lambda) = EX^r = \frac{\Gamma(r+1)}{\lambda^r}.$$

(b) Use the moment calculation in (a) to show that

$$\frac{\mu_r(\lambda)}{\mu_{r+1}(\lambda)} = \frac{\lambda}{r+1}$$

and hence that the family of estimators $\{\hat{\lambda}_n^{(k)}\}_{k\geq 0}$ given by

$$\hat{\lambda}_n^{(k)} \equiv (k+1)\frac{\overline{X^k}_n}{\overline{X^{k+1}}_n} \equiv (k+1)\frac{n^{-1}\sum_1^n X_i^k}{n^{-1}\sum_1^n X_i^{k+1}}$$

are all consistent estimators of λ: $\hat{\lambda}_n^{(k)} \to_p \lambda$ for each $k = 0, 1, 2, \ldots$.
(c) Show that

$$\sqrt{n}(\hat{\lambda}_n^{(k)} - \lambda) \to_d N(0, \sigma_k^2(\lambda)) \ \text{ as } \ n \to \infty$$

and compute $\sigma_k^2(\lambda)$ explicitly as a function of k and λ.
(d) What is the asymptotic relative efficiency of $\hat{\lambda}_n^{(k)}$ to $\widehat{\lambda}_n \equiv \hat{\lambda}_n^{(0)} = 1/\overline{X}_n$ for $k > 1$?
(e) Now suppose that $X, X_1, \ldots, X_n$ are i.i.d. with distribution function F on $(0, \infty)$ where F is not an exponential distribution function. Specify hypotheses on F (or X) which guarantee that $\hat{\lambda}_n^{(k)} \to_p$ some natural parameter, say $\lambda_k(F)$ defined in terms of F. What hypothesis will be needed to guarantee that $\sqrt{n}(\hat{\lambda}_n^{(k)} - \lambda_k(F)) \to_d N(0, V^2)$ for some V^2?

Solution:
(a) We compute

$$
\begin{aligned}
E(X^r) &= \int_0^\infty x^r \lambda e^{-\lambda x} dx = \lambda^{-r} \int_0^\infty (\lambda x)^r e^{-\lambda x} \lambda dx \\
&= \lambda^{-r} \int_0^\infty y^{(r+1)-1} e^{-y} dy = \lambda^{-r} \Gamma(r+1).
\end{aligned}
$$

(b) It follows from (a) that

$$\frac{\mu_r(\lambda)}{\mu_{r+1}(\lambda)} = \frac{\lambda}{r+1}$$

and hence

$$
\begin{aligned}
\hat{\lambda}_n^{(k)} &\equiv (k+1)\frac{\overline{X^k}_n}{\overline{X^{k+1}}_n} \equiv (k+1)\frac{n^{-1}\sum_1^n X_i^k}{n^{-1}\sum_1^n X_i^{k+1}} \\
&\to_p (k+1)\frac{\mu_k(\lambda)}{\mu_{k+1}(\lambda)} = \lambda.
\end{aligned}
$$

(c) Now by the multivariate CLT it follows that

$$\sqrt{n}\begin{pmatrix} \overline{X^k}_n - \mu_k \\ \overline{X^{k+1}}_n - \mu_{k+1} \end{pmatrix} \to_d \underline{Z} \sim N_2(0, \Sigma)$$

where

$$
\begin{aligned}
\Sigma &= \begin{pmatrix} \frac{\Gamma(2k+1)-\Gamma(k+1)^2}{\lambda^{2k}} & \frac{\Gamma(2k+2)-\Gamma(k+1)\Gamma(k+2)}{\lambda^{2k+1}} \\ \frac{\Gamma(2k+2)-\Gamma(k+1)\Gamma(k+2)}{\lambda^{2k+1}} & \frac{\Gamma(2k+3)-(\Gamma(k+2)^2}{\lambda^{2k+2}} \end{pmatrix} \\
&= \frac{1}{\lambda^{2k}}\begin{pmatrix} \Gamma(2k+1)-\Gamma(k+1)^2 & \frac{\Gamma(2k+2)-\Gamma(k+1)\Gamma(k+2)}{\lambda^1} \\ \frac{\Gamma(2k+2)-\Gamma(k+1)\Gamma(k+2)}{\lambda^1} & \frac{\Gamma(2k+3)-(\Gamma(k+2)^2}{\lambda^2} \end{pmatrix}.
\end{aligned}
$$

Thus by the delta method with $g(u,v) = u/v$, so that $\dot{g}(u,v) = v^{-1}(1, -u/v)$

$$
\begin{aligned}
\sqrt{n}(\hat{\lambda}_n^{(k)} - \lambda) &= (k+1)\sqrt{n}(g(\overline{X^k}_n, \overline{X^{k+1}}_n) - g(\mu_k(\lambda), \mu_{k+1}(\lambda)) \\
&\to_d (k+1)\dot{g}(\mu_k(\lambda), \mu_{k+1}(\lambda))\underline{Z} \\
&= \frac{k+1}{\mu_{k+1}(\lambda)}\left(Z_1 - \frac{\mu_k}{\mu_{k+1}}Z_2\right) \\
&= \frac{1}{\mu_{k+1}}((k+1)Z_1 - \lambda Z_2) \\
&\sim \frac{1}{\mu_{k+1}}N(0, \lambda^{-2k}C_k) = N(0, \frac{1}{\lambda^{2k}\mu_{k+1}^2}C_k) \\
&= N(0, \lambda^2\frac{C_k}{\Gamma(k+2)^2}) \equiv N(0, \lambda^2 D_k)
\end{aligned}
$$

where

$$C_k = (k+1)^2 \left\{\Gamma(2k+1) - \Gamma(k+1)^2\right\} - 2(k+1)\{\Gamma(2k+2) - \Gamma(k+1)\Gamma(k+2)\} + \Gamma(2k+3) - \Gamma(k+2)^2.$$

and (after a bit of algebra)

$$D_k = \frac{\Gamma(2k+1)}{\Gamma(k+1)^2}\left\{1 - 2\frac{2k+1}{k+1} + \frac{(2k+2)(2k+1)}{(k+1)^2}\right\} = \frac{\Gamma(2k+1)}{\Gamma(k+1)^2}.$$

(c) When $k = 0$, we compute $D_k = 1$. Thus the asymptotic relative efficiency of $\hat{\lambda}_n^{(k)}$ with respect to $\hat{\lambda}_n^{(0)}$ is D_0/D_k. These estimators become inefficient relative to the mean very rapidly as k increases, as is shown by the following plot of the relative efficiency.

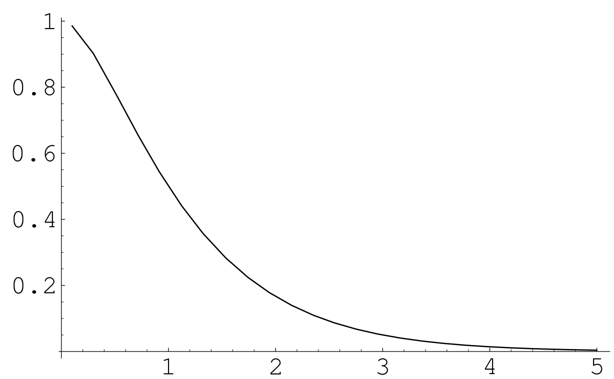

Figure 5: Asymptotic relative efficiency of $\hat{\lambda}_n^{(k)}$ with respect to $\hat{\lambda}_n^{(0)}$